



Universitat d'Alacant
Universidad de Alicante

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant

Una aproximació semàntica al processament de la informació temporal

Héctor Llorens Martínez

Tesi Doctoral

Aquest treball s'ha dut a terme sota la supervisió de
la Dra. Estela Saquete Boró i el Dr. Borja Navarro-Colorado,
Universitat d'Alacant

Alacant, juliol 2011



Universitat d'Alacant
Universidad de Alicante

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant

Una aproximació semàntica al processament de la informació temporal

Héctor Llorens Martínez

Tesi Doctoral

Aquest treball s'ha dut a terme sota la supervisió de
la Dra. Estela Saquete Boró i el Dr. Borja Navarro-Colorado,
Universitat d'Alacant

Alacant, juliol 2011

Aquesta tesi doctoral ha sigut finançada pel Govern espanyol, en els projectes: TIN-2006-15265-C06-01, TIN-2009-13391-C04-01 i PROMETEO/2009/119, on Héctor Llorens gaudeix d'una beca FPI (BES-2007-16256).

“El temps... és el que evita que tot passe alhora”

Ray Cummings (1922)

Agraïments

Montse, gràcies per fer de la meua vida un somni tan plaent i per haver-me acompanyat i animat durant tot el procés predoctoral.

Andrés, Emília, David, Rosa, gràcies per estar sempre quan vos necessite.

Xavi, company de batalla, tinc sort que hàguem seguit el mateix camí acadèmic, mai no m'he sentit sol.

Estela, Borja, ha sigut gratificant fer aquesta tesi sota la vostra guia.

Agraïsc al grup i a totes les persones que m'han acompanyat en aquest procés les bones experiències compartides.

Pàgina dedicada als que he estimat i ja no estan.

Resum

Aquesta tesi se centra en el processament de la informació temporal, que és una tasca emmarcada en el camp del processament del llenguatge natural. L'objectiu d'aquesta tasca és obtenir la representació i ordenació temporal dels esdeveniments expressats en el text o discurs, la qual cosa requereix la interpretació automàtica de les expressions temporals, els esdeveniments, i les seues relacions temporals.

La majoria de les aproximacions actuals estan basades en l'ús de la informació morfosintàctica. No obstant això, les entitats temporals són, sovint, ambigües a aquest nivell d'anàlisi lingüística.

La nostra hipòtesi és que l'expressió lingüística del temps és un fenomen semàntic i, per tant, per a obtenir-ne un millor rendiment, la informació temporal ha de ser processada usant també informació semàntica.

Per a provar aquesta hipòtesi, presentem una aproximació semàntica al processament de la informació temporal: TIPSem. Es tracta d'un sistema automàtic que inclou característiques basades en semàntica lèxica i rols semàntics, a més de característiques morfosintàctiques.

TIPSem ha sigut avaluat mitjançant la seua participació en l'exercici internacional d'avaluació TempEval-2. Els resultats obtinguts avalen la hipòtesi presentada, i l'anàlisi detallada d'aquests demostra que les característiques semàntiques afavoreixen les capacitats de generalització del model i ajuden a gestionar l'ambigüitat morfosintàctica. Aquestes conclusions s'han assolit per a diferents llengües (*i. e.*, anglès, castellà, italià, i xinès), la qual cosa avala la hipòtesi a escala multilingüe. Finalment, en comparació amb l'estat de la qüestió, TIPSem ofereix un rendiment molt competitiu i introdueix una millora notable en el processament d'esdeveniments.

A més, TIPSem ha sigut aplicat al problema de la representació gràfica de la informació temporal. S'ha desenvolupat una interfície dinàmica que ofereix als usuaris un accés a la informació basat en el temps: Time-Surfer. Els resultats obtinguts mitjançant una avaluació orientada a l'usuari d'aquesta interfície demostren que el rendiment de TIPSem és també satisfactori des d'un punt de vista extrínsec.

Índex

| | | |
|----------|--|-----------|
| 1 | Introducció | 1 |
| 1.1 | Objectius | 4 |
| 1.2 | Estructura de la tesi | 5 |
| 2 | Estat de la qüestió | 7 |
| 2.1 | Fonaments | 8 |
| 2.2 | Esquemes d'anotació temporal | 19 |
| 2.3 | Aproximacions computacionals actuals | 30 |
| 2.4 | L'ús de la semàntica | 36 |
| 2.4.1 | Semàntica lèxica | 36 |
| 2.4.2 | Rols semàntics | 37 |
| 2.5 | Conclusions | 40 |
| 3 | TIPSem: Una aproximació semàntica | 41 |
| 3.1 | Arquitectura general | 42 |
| 3.2 | Arquitectura específica de cada tasca | 44 |
| 3.2.1 | Processament de <i>timex</i> | 45 |
| 3.2.2 | Processament d'esdeveniments | 55 |
| 3.2.3 | Processament de relacions temporals | 62 |
| 3.3 | Selecció de tècniques d'aprenentatge automàtic | 74 |
| 3.4 | Extensió al castellà, italià i xinès | 77 |
| 3.5 | Conclusions | 82 |
| 4 | Avaluació | 87 |
| 4.1 | Avaluació sobre text en anglès | 87 |
| 4.2 | Avaluació multilingüe | 118 |
| 4.3 | Conclusions | 136 |

ÍNDIX

| | | |
|----------|--|------------|
| 5 | Aplicació: Time-Surfer | 139 |
| 5.1 | Representant el temps gràficament | 140 |
| 5.2 | Treballs relacionats | 143 |
| 5.3 | Time-Surfer: arquitectura i característiques | 143 |
| 5.4 | Avaluació de Time-Surfer | 146 |
| 5.5 | Conclusions | 153 |
| 6 | Conclusions | 155 |
| 6.1 | Contribució | 155 |
| 6.2 | Treball futur | 159 |
| 6.3 | Producció científica | 162 |
| 6.4 | Producció de programari | 164 |
| A | TIMEK: Base de coneixements temporals | 167 |
| B | Comparació de resultats CRF-SVM | 173 |
| C | Validació creuada i rellevància estadística | 177 |
| D | Exercicis d'avaluació TempEval | 193 |
| | Bibliografia | 197 |

Índex de taules

| | | |
|------|---|-----|
| 2.1 | Relacions temporals dels temps verbals (* ambigües) | 11 |
| 2.2 | Propietats de les categories aspectuals | 12 |
| 2.3 | Comparativa del esquemes d' anotació temporal | 29 |
| 2.4 | Aproximacions de processament TimeML per a l' anglès | 34 |
| 3.1 | Selecció de tècniques d' ML | 76 |
| 4.1 | Estadística de les dades TimeML en anglès (TempEval-2) . . | 89 |
| 4.2 | Acord entre anotadors en TimeML anglès | 90 |
| 4.3 | P&R i Kappa del TimeML anglès | 90 |
| 4.4 | Reconeixement de <i>timex</i> (anglès) | 93 |
| 4.5 | Classificació de <i>timex</i> (anglès) | 97 |
| 4.6 | Normalització de <i>timex</i> (anglès) | 98 |
| 4.7 | Reconeixement d'esdeveniments (anglès) | 102 |
| 4.8 | Reconeixement d'esdeveniments (detall per categoria gramatical) | 104 |
| 4.9 | Classificació d'esdeveniments (anglès) | 105 |
| 4.10 | Classificació d'esdeveniments (detall per classe gramatical) . . | 106 |
| 4.11 | Classificació d'esdeveniments (detall per classe) | 107 |
| 4.12 | Categorització de relacions event- <i>timex</i> (anglès) | 108 |
| 4.13 | Resultats de categorització de relacions event-DCT (anglès) . | 111 |
| 4.14 | Categorització de relacions main-event (anglès) | 113 |
| 4.15 | Categorització de subord-event (anglès) | 114 |
| 4.16 | Resum de resultats per a l' anglès | 116 |
| 4.17 | Estadístiques de les dades TimeML en castellà (TempEval-2) | 119 |
| 4.18 | Reconeixement de <i>timex</i> (espanyol) | 119 |
| 4.19 | Classificació de <i>timex</i> (castellà) | 120 |
| 4.20 | Normalització de <i>timex</i> (castellà) | 121 |
| 4.21 | Reconeixement d'esdeveniments (castellà) | 123 |

ÍNDIX DE TAULES

| | | |
|------|--|-----|
| 4.22 | Classificació d'esdeveniments (castellà) | 123 |
| 4.23 | Categorització de relacions event- <i>timex</i> (castellà) | 124 |
| 4.24 | Categorització de relacions Event-DCT (castellà) | 125 |
| 4.25 | Resum de resultats per al castellà | 127 |
| 4.26 | Dades en italià (esdeveniments TempEval-2) | 129 |
| 4.27 | Reconeixement d'esdeveniments (italià) | 129 |
| 4.28 | Classificació d'esdeveniments (italià) | 131 |
| 4.29 | Estadístiques de les dades en xinès (TempEval-2) | 133 |
| 4.30 | Reconeixement de <i>timex</i> (xinès) | 133 |
| 4.31 | Reconeixement d'esdeveniments (xinès) | 134 |
| 4.32 | Resum dels resultats de TIPSem per a l'anglès (MRER_S: reducció mitjana de l'error relatiu introduït per la semàntica, SSC: Grau de confiança en la millora estadísticament significativa) | 137 |
| B.1 | Processament de <i>timex</i> CRF vs. SVM (anglès) | 174 |
| B.2 | Processament d'esdeveniments CRF vs. SVM (anglès) | 174 |
| B.3 | Processament de relacions temporals CRF vs. SVM (anglès) | 174 |
| B.4 | Selecció de tècniques d'ML en TIPSem | 175 |
| C.1 | 10 conjunts de dades (entitats anglès) | 178 |
| C.2 | 10 conjunts de dades (relacions anglès) | 178 |
| C.3 | Reconeixement de <i>timex</i> 10-fold (anglès) | 182 |
| C.4 | Classificació de <i>timex</i> 10-fold (anglès) | 183 |
| C.5 | Normalització de <i>timex</i> 10-fold (anglès) | 184 |
| C.6 | Reconeixement d'esdeveniments 10-fold (anglès) | 185 |
| C.7 | Classificació d'esdeveniments 10-fold (anglès) | 186 |
| C.8 | Categorització Event- <i>timex</i> 10-fold (anglès) | 187 |
| C.9 | Categorització Event-DCT 10-fold (anglès) | 188 |
| C.10 | Categorització d'esdeveniments principals 10-fold (anglès) | 189 |
| C.11 | Categorització d'esdeveniments subordinats 10-fold (anglès) | 190 |
| C.12 | Resum de la validació de 10 conjunts (RER_S: reducció relativa d'error per la semàntica, SC: fiabilitat de la significació) | 191 |
| D.1 | TempEval-1 resultats F1 | 194 |
| D.2 | TempEval-2 - resultats de processament de <i>timex</i> | 195 |
| D.3 | TempEval-2 - resultats de processament d'esdeveniments | 196 |
| D.4 | TempEval-2 - resultats de processament de relacions | 196 |

Índex de figures

| | | |
|------|--|-----|
| 1.1 | Escenari de processament de la informació temporal | 2 |
| 1.2 | La necessitat de la semàntica en el processament d'informació temporal | 3 |
| 2.1 | Estructura (nucli) de l'esdeveniment basat en la contingència | 13 |
| 2.2 | Xarxa de transaccions per al càlcul aspectual | 13 |
| 2.3 | Les 13 relacions temporals d'Allen | 15 |
| 2.4 | Relacions temporals | 16 |
| 2.5 | Xarxa de relacions temporals | 16 |
| 2.6 | Part de la taula de transitivitat d'Allen | 17 |
| 2.7 | Descomposició de una relació entre intervals en relacions entre punts | 17 |
| 3.1 | Arquitectura del TIPSem | 42 |
| 3.2 | TIPSem arquitectura de processament de <i>timex</i> | 45 |
| 3.3 | TIPSem característiques per al reconeixement de <i>timex</i> . . . | 51 |
| 3.4 | Arquitectura de processament d'esdeveniments del TIPSem . | 56 |
| 3.5 | Característiques del TIPSem per al reconeixement d'esdeveniments | 60 |
| 3.6 | Arquitectura de processament de relacions temporals de TIPSem | 65 |
| 3.7 | Característiques de TIPSem per a event- <i>timex</i> | 68 |
| 3.8 | Característiques de TIPSem per a event-DCT | 69 |
| 3.9 | Característiques del TIPSem per a relacions <i>main-event</i> . . . | 71 |
| 3.10 | Característiques de TIPSem per a <i>subordinated-event</i> | 73 |
| 3.11 | Camps Aleatoris Condicionals | 74 |
| 3.12 | SVM | 75 |
| 3.13 | Verbs del lèxic SIMPLE/CLIPS | 79 |
| 4.1 | Mesurant la contribució de la semàntica | 88 |
| 4.2 | Comparativa de resultats amb l'estat de la qüestió (anglès) . | 117 |

ÍNDIX DE FIGURES

| | | |
|------|---|-----|
| 4.3 | Comparativa amb l'estat de la qüestió (castellà) | 128 |
| 5.1 | Problema de la representació de línia de temps I | 141 |
| 5.2 | Problema de la representació de línia de temps II | 142 |
| 5.3 | Representació amb graf temporal | 142 |
| 5.4 | Arquitectura Time-Surfer | 144 |
| 5.5 | Buscant <i>Alemanya</i> en WWI, assenyalant 1939, i fent zoom . | 146 |
| 5.6 | Comparació de WWI i Einstein | 146 |
| 5.7 | Time-Surfer <i>vs.</i> Viquipèdia (mitjana de científics de la com- putació) | 150 |
| 5.8 | Time-Surfer satisfacció (científics de la computació) | 151 |
| 5.9 | Time-Surfer <i>vs.</i> Viquipèdia (Filologia Hispànica) | 152 |
| 5.10 | Satisfacció sobre l'ús de Time-Surfer (Filologia Hispànica) . . | 153 |

Capítol 1

Introducció

L'enorme quantitat d'informació digitalitzada i la seua àmplia difusió a través d'Internet ha anomenat l'època actual com la *Societat de la Informació* (Beniger, 1986). No obstant això, la quantitat d'informació ha sobrepassat els límits humans. En aquest context, el repte és millorar o automatitzar l'accés a la informació i la seua gestió per a aprofitar aquesta font tan valuosa de coneixements.

En l'actualitat, els ordinadors només poden processar informació estructurada o semi-estructurada on s'esmenta o codifica explícitament el significat, com en una base de dades o un arxiu XML. No obstant això, en aquest moment, la majoria de la informació disponible no està estructurada – és principalment text en llenguatge natural. El repte és, per tant, aconseguir que els ordinadors puguin processar la informació en llenguatge natural, convertint-la en informació estructurada.

Aquesta qüestió s'aborda des de l'àrea d'investigació d'intel·ligència artificial, i més concretament de processament del llenguatge natural (PLN). Dins del PLN, aquesta tesi aborda el processament computacional de la informació temporal expressada en llenguatge natural. Aquesta tasca implica la localització i ordenació dels esdeveniments en el temps, la qual cosa exigeix la conversió automàtica d'expressions temporals, esdeveniments, i les seues relacions en informació estructurada per al seu posterior ús en aplicacions més complexes.

La figura 1.1 il·lustra la motivació per al tractament de la informació temporal. Aquesta figura mostra un text que parla d'estrenes de pel·lícules, la qual cosa representa un conjunt d'esdeveniments situats en el temps. En aquest context, l'objectiu és obtenir una representació estructurada que relaciona les pel·lícules amb les dates del seu llançament.

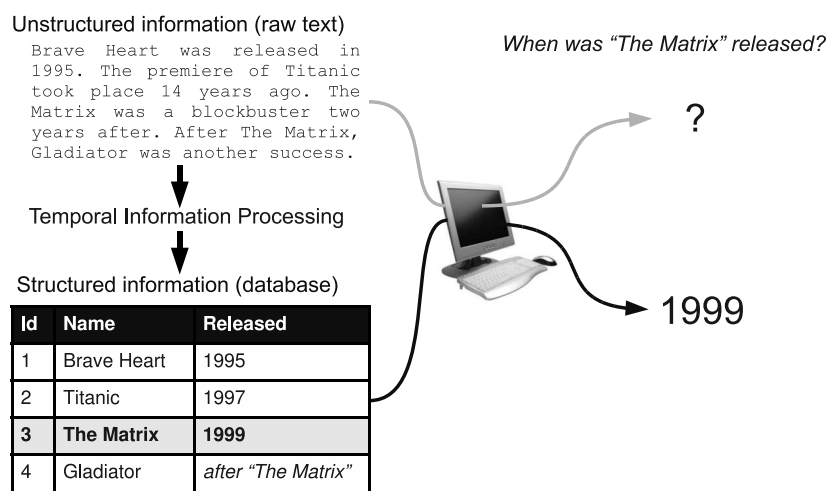


Figura 1.1: Escenari de processament de la informació temporal

La complexitat de la tasca radica en l'ambigüïtat del llenguatge natural. Els esdeveniments, les referències de temps, i les seues relacions a voltes són explícites, com en *Brave Heart was released in 1995*, però no sempre ho són. Poden ser implícites, com en *blockbuster* i *two years after*. O fins i tot alguna informació temporal pot ser només parcialment especificada com en *Gladiator*, on la data de llançament s'elideix i només és possible ordenar l'esdeveniment relativament *after The Matrix*.

Com es pot observar, la complexitat de la tasca va des del reconeixement de patrons simples (per exemple, [pel·lícula] [publicat en] [data]) fins a la comprensió del llenguatge natural per a inferir que, per exemple, *blockbuster* implica un llançament de la pel·lícula i *two years after Titanic* és refereix a la data de llançament de *Titanic* més 2 anys.

La importància i l'interès de la investigació en el processament de la informació temporal recau sobre l'avantatge que moltes aplicacions del PLN prenen de tenir en compte la informació temporal (Mani et al., 2005; Schilder et al., 2007; Verhagen et al., 2010). Per exemple, els sistemes de cerca de erspostes requereixen raonament temporal per a respondre qüestions temporals, com *Qui era el president d'EUA fa deu anys?* i *Què va passar amb els preus del petroli del món després de l'annexió de Kuwait per l'Iraq?* (Pustejovsky, 2002; Saquete et al., 2009). En el resum de text automàtic, la capacitat d'assignar als esdeveniments les dates en les quals van ocórrer ajuda a millorar resums quan aquests han de ser enfocats a un

període de temps determinat, o sobre esdeveniments concrets (Daniel et al., 2003). En la recuperació de la informació i la visualització, el valor de la dimensió temporal ha estat subratllada en la investigació (Alonso et al., 2007a, 2010; Verhagen, 2007) i en aplicacions reals (per exemple, Google timeline, Yahoo Time-Explorer).

L'èxit de les aplicacions abans esmentades depèn directament de l'eficàcia de l'aproximació de processament de la informació temporal emprat. Un error del processament temporal impedeix que l'aplicació final pugui obtenir la interpretació temporal correcta del text, com es mostra en la figura 1.2.

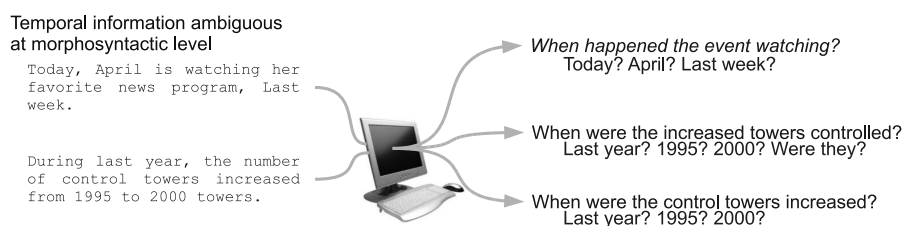


Figura 1.2: La necessitat de la semàntica en el processament d'informació temporal

La semàntica és necessària per a interpretar correctament la informació temporal expressada en llenguatge natural. L'exemple (1) mostra algunes frases en les quals el component temporal és ambigu en els nivells inferiors d'anàlisi del llenguatge.

- (1) a. April is watching her favorite news program, Last Week.
 b. The number of control towers increased from 1995 to 2000 towers.
 c. Currently, they control the situation that was chaotic from 1995 to 2000.
 d. A UN weapons inspection team resumed work.
 e. There was an inspection in our department.
 f. Is is a good work.
 g. I went back home two hours ago, after finishing my research.
 h. I went back home two hours ago, and now I'm finishing my research.

Encara que *April* i *Last Week* són normalment expressions temporals, no juguen un paper temporal en (1a), sinó que representen un nom propi femení i el nom d'un noticiari. Les frases (1b) i (1c) mostren moltes similituds morfosintàctiques, però, *1995* i *2000* només són expressions temporals en (1c). També, el lema *control* és només un esdeveniment en (1c). En la frase (1d), *work* es un esdeveniment i *inspection* no ho és. No obstant això, en

(1e), *inspection* es un esdeveniment, i en (1f) *inspection* no ho és. Pel que es refereix a les relacions temporals, en (1g), *finishing* passa després de *went* mentre que és al contrari en (1h).

En aquests casos, el coneixement sobre la semàntica de les paraules i les seues relacions a nivell de predicat són crucials per a determinar quines paraules denoten expressions temporals o esdeveniments, i també quin tipus de relacions temporals representen. Aquesta és, precisament, la nostra motivació per a aplicar la semàntica, a més de la morfosintaxi, al processament de la informació temporal.

La nostra hipòtesi és que amb la semàntica els resultats dels enfocaments actuals es poden millorar. Plantegem que els models construïts utilitzant la semàntica seran més generals que els basats únicament en els elements morfosintàctics a causa del nivell d'abstracció que suposa la informació semàntica sobre les paraules i categories gramaticals (PoS¹); i també a causa de l'ambigüïtat morfosintàctica. D'altra banda, la contribució de la semàntica ha de ser comparable en diferents idiomes, ja que és un nivell d'anàlisi del llenguatge alt.

1.1 Objectius

La hipòtesi anterior s'ha pres com a base per a definir els objectius d'aquesta tesi. Els objectius específics i les línies d'actuació proposades per a aconseguir-los s'enumeren a continuació.

- **Definir una proposta:** Proposem definir i implementar un model i una metodologia per a l'aplicació de la semàntica al processament de la informació temporal en diferents idiomes. Per a açò, durem a terme un estudi dels diferents tipus de semàntica centrat en els beneficis que aquests poden oferir en la tasca. Prenent com a base aquest estudi, definirem i aplicarem una aproximació semàntica per al processament de la informació temporal en anglès (**TIPSem**), que inclou característiques semàntiques, a més de les característiques morfosintàctiques. Finalment, aquest plantejament s'ampliarà a l'espanyol, l'italià i el xinès.
- **Avaluar la proposta:** Proposem una avaluació intrínseca de la proposta per a mesurar empíricament la influència dels trets semàntics en la tasca del processament de la informació temporal per a diferents

¹En anglès, Part-of-Speech (PoS).

idiomes. Per a açò, es desenvoluparà una aproximació *baseline* centrada només en les característiques morfosintàctiques (TIPSem-B), i serà avaluat juntament amb TIPSem sobre el corpus de l'exercici d'avaluació internacional TempEval-2², la qual cosa afavoreix la possibilitat de reproduir els experiments i permet a altres investigadors de comparar altres enfocaments amb els nostres. D'altra banda, els resultats seran analitzats centrant-nos en la contribució de les diferents característiques semàntiques tant de forma independent com en combinació. Finalment, l'avaluació s'estendrà a espanyol, italià i xinès per a analitzar els avantatges dels trets semàntics i el seu grau d'independència del llenguatge en un entorn multilingüe.

- **Aplicar i avaluar la proposta extrínsecament:** Proponem una avaluació extrínseca de la proposta a través de la seua aplicació. Per a això, desenvoluparem una interfície gràfica basada en el temps (**Time-Surfer**), en el qual TIPSem s'aplicarà com una peça fonamental del procés, i realitzarem una avaluació orientada a l'usuari d'aquesta interfície per a avaluar el rendiment del TIPSem extrínsecament.

1.2 Estructura de la tesi

La tesi està estructurada de la manera següent.

El capítol 2 revisa els treballs relacionats rellevants per a aquesta tesi. Això inclou els fonaments del processament d'informació temporal i la revisió dels enfocaments actuals. A més, s'analitzen els recursos semàntics que s'apliquen en aquesta tesi.

El capítol 3 descriu el nostre enfocament semàntic al processament de la informació temporal: **TIPSem**. En aquest capítol s'explica cada part de l'arquitectura del sistema, especialment els trets semàntics que s'han definit.

El capítol 4 inclou l'avaluació de la proposta sobre el conjunt de dades TempEval-2 per a l'anglès. Els resultats obtinguts pel nostre enfocament semàntic es comparen amb un *baseline* morfosintàctic i amb l'estat de la qüestió analitzant els avantatges introduïts pels trets semàntics. D'altra banda, s'inclou una avaluació multilingüe que analitza la contribució de la semàntica en espanyol, italià i xinès.

²<http://www.timeml.org/tempeval2/>

El capítol 5 descriu l'aplicació del TIPSem. Es defineix una interfície gràfica basada en el temps: **Time-Surfer**, en la qual TIPSem s'utilitza com un component crucial. Aquest capítol també inclou una avaluació orientada cap als usuaris d'aquesta aplicació, en el qual s'analitzen els resultats per a avaluar el rendiment del TIPSem extrínsecament.

El capítol 6 llista les conclusions i també les direccions futures d'aquesta investigació.

Aquesta tesi ha sigut finançada pel govern d'Espanya, en els projectes: TIN-2006-15265-C06-01, TIN-2009-13391-C04-01 i PROMETEO/2009/119, on H. Llorens gaudeix d'una beca FPI (BES-2007-16256). En el marc d'aquests projectes, aquesta tesi aborda l'objectiu de desenvolupar tècniques i eines per a extraure informació temporal dels textos.

Capítol 2

Estat de la qüestió

Com a part essencial de la intel·ligència artificial (IA), el processament del llenguatge natural (PLN) investiga algorismes computacionals eficaços capaços d'analitzar, comprendre i generar llenguatge natural parlat, escrit o signat (Moreno et al., 1999; Allen, 1995; Jurafsky & Martin, 2000).

El processament de la informació temporal és una tasca del PLN que se centra en la interpretació automàtica dels aspectes temporals del llenguatge natural. Els éssers humans comprenen el món d'una manera que està estretament lligada a la percepció del temps. Per aquesta raó, els llenguatges naturals inclouen mecanismes per a expressar el temps (Lyons, 1968), com ara temps verbal, l'aspecte i la modalitat (per exemple, futur, passat perfecte), els adverbis i els sintagmes temporals (per exemple, *ahir*, *d'octubre*), també preposicions i conjuncions específiques (per exemple, *abans*, *quan*), etc¹. Mitjançant l'ús d'aquests mecanismes, podem expressar les coses que succeeixen (esdeveniments), les seues propietats temporals (estats), i la seua localització en el temps. Això fa que siga possible recrear en la ment del receptor una visió temporal del que s'expressa. L'exemple (2) mostra un text on la component temporal és important.

- (2) Ell va arribar a la festa a les 20:00 h.
No obstant això, ella ja se n'havia anat.
Ell tornà a casa, després de parlar amb alguns amics.

A partir d'aquestes frases, el lector pot recrear la realitat següent. Hi ha un esdeveniment (*arribar*), que va tenir lloc en un punt del temps específic (*20:00*), i un altre esdeveniment (*anar-se'n*) que va passar abans

¹Això varia per als diferents idiomes. Alguns idiomes, amb diferències notables són el xinès (Li et al., 2001), hopi (Whorf, 1956), nootka, i birmanès (Comrie, 1985).

que el primer. Aquests esdeveniments produeixen canvis d'estat en els seus participants (*ell* i *ella*) per a estar o no en la festa. Un esdeveniment posterior (*tornà*) té lloc després d'un esdeveniment relacionat (*parlar*), i ambdós succeeixen després de l'esdeveniment *arribar*.

El processament de la informació temporal automatitza la tasca duta a terme en l'exemple anterior. Aquesta tasca implica la identificació de les entitats temporal en el llenguatge natural i la interpretació de les seues propietats temporals. A més, les seues relacions han de ser interpretades amb la finalitat d'ubicar i ordenar els successos en el temps. Aquesta és una tasca complexa que requereix coneixements lingüístics en tots els nivells d'anàlisi del llenguatge, incloent-hi el semàntic, el pragmàtic i el nivell de discurs.

Abans de descriure la forma en que el processament d'informació s'aborda des de la perspectiva del PLN, s'introdueixen els estudis preliminars sobre el temps i el llenguatge; en segon lloc, ens dirigim a l'actual metodologia del PLN que es basa en la lingüística de corpus i els esquemes d'anotació; en tercer lloc, es revisa l'estat de la qüestió de les aproximacions computacionals; i finalment, es presenten els diferents tipus de semàntica aplicada en l'àmbit d'aquesta tesi.

2.1 Fonaments

Per a comprendre millor els enfocaments actuals de processament de la informació temporal, cal revisar els antecedents o fonaments. Les subseccions següents revisen les obres més destacades sobre el temps i el llenguatge des de tres perspectives diferents (Mani et al., 2005). El primer incís es refereix a les teories lingüístiques de temps, aspecte i modalitat. El segon apartat revisa el raonament temporal des de la perspectiva de la IA. Finalment, el tercer apartat analitza les obres de l'estructura temporal del discurs.

Temps verbal, aspecte, i modalitat

L'expressió gramaticalitzada del temps i les propietats temporals dels esdeveniments es realitza en el llenguatge a través de temps verbal, l'aspecte i la modalitat (Lyons, 1981). Per tant, l'estudi d'aquests elements és rellevant per al desenvolupament de mètodes computacionals de processament de la

informació temporal.

Temps verbal

El temps dels verbs és un mecanisme específic per a la localització en el temps dels fets que aquests expressen (Comrie, 1985; Crystal, 2003). L'exemple (3) mostra com els diferents temps localitzen un esdeveniment en el passat, en el present o en el futur pel que fa al temps de parla.

- (3) a. John menjava (passat).
b. John menja (present).
c. John menjarà (futur).

Reichenbach (1947) va presentar un model temporal de la representació dels temps verbals que consta de tres punts de temps. Ell va introduir el punt *temps de referència*.

- **Temps de conversa o escriptura (S)**: Moment de la descripció lingüística de l'esdeveniment.
- **Temps de referència (R)**: Temps de què es parla.
- **Temps de l'esdeveniment (E)**: Moment en el qual l'esdeveniment descrit ocorre.

Els temps estan representats sobre aquests punts amb els operadors de *precedeix* ($<$) i *coincideix* ($=$). Per exemple, el passat simple (per exemple, *guanyava*), representa una referència de temps anterior al temps de conversa en què l'esdeveniment *guanyar* ocorre ($E = R < S$). No obstant això, en el passat perfecte (per exemple, *havia guanyat*), l'esdeveniment *guanyar* es produeix abans d'un temps de referència, que és anterior al temps de conversa ($E < R < S$). Sovint la referència temporal explícita s'elideix, però es pot representar mitjançant expressions temporals com *ahir* (per exemple, *guanyà ahir*).

Reichenbach defineix un conjunt de relacions entre aquests punts per representar tots els temps. Atès que algunes relacions de Reichenbach eren ambigües (Comrie, 1985), es va desenvolupar un conjunt inequívoc de les relacions afegint l'operador *segueix* ($>$) i presentant les relacions sempre en l'ordre SRE (Song & Cohen, 1991).

La taula 2.1 il·lustra les relacions originals de Reichenbach, la variant sense ambigüitats, i la seua correspondència als temps verbals en anglès.

El treball de Reichenbach ha influït en el treball posterior en la lingüística formal i computacional, especialment l'anàlisi de l'estructura temporal del discurs (Webber, 1988).

| Reichenbach relations | unambiguous relations | English tense | example |
|---------------------------|-----------------------|-----------------|----------------------------------|
| E<R<S (Anterior Past) | S>R>E | Past Perfect | <i>I had slept</i> |
| E=R<S (Simple Past) | S>R=E | Past Simple | <i>I slept</i> |
| R<E<S* (Posterior Past) | S>R<E | | <i>[I expect] I would sleep</i> |
| R<S=E* | | | |
| R<S<E* | | | |
| E<S=R (Anterior Present) | S=R>E | Present Perfect | <i>I have slept</i> |
| S=R=E (Simple Present) | S=R=E | Present Simple | <i>I sleep</i> |
| S=R<E (Posterior Present) | S=R<E | Future Simple | <i>I will sleep</i> |
| S=E<R* | | | |
| E<S<R* | | | |
| S<E<R* (Anterior Future) | S<R>E | Future Perfect | <i>I will have slept</i> |
| S<R=E (Simple Future) | S<R=E | Future Simple | <i>I will sleep</i> |
| S<R<E (Posterior Future) | S<R<E | | <i>I shall be going to sleep</i> |

Taula 2.1: Relacions temporals dels temps verbals (* ambigües)

Aspecte

En aquest apartat, podem distingir entre l'aspecte lèxic i l'aspecte gramatical.

L'aspecte lèxic recull les següents propietats temporals dels esdeveniments: dinamisme², acabament, i duració (Vendler, 1967; Dowty, 1979). Pel que fa a aquestes propietats, els esdeveniments poden ser classificats en les següents categories aspectuals o *Aktionsarten*:

- **Estats:** Aquests representen les propietats estàtiques de les entitats que es mantenen durant un interval de temps, i que poden canviar en un instant determinat en el temps, com podem veure en (4).

- (4) a. M'agrada el futbol.
b. L'ordinador està trencat.
c. Ella estava trista, però ara està contenta.

- **Esdeveniments**³: A diferència dels estats, aquests descriuen situacions dinàmiques. De vegades, els esdeveniments representen un canvi – canvi de la situació actual de la realitat. Per exemple, en (4b), l'esdeveniment *reparar* implicaria un canvi d'estat de *trencat* a *reparat*. D'acord amb Vendler, un esdeveniment pertany a una de les següents classes en funció de les seues propietats temporals:

²També es coneix com a eventualitat.

³*events* (en anglès) es un terme general ambigu per a referir-se a esdeveniments i estats. De vegades *eventualitats* (1986) s'empra com a terme general no ambigu.

- **Processos duratius no acabats:** Esdeveniments que impliquen una duració, però sense un punt final definit. Alguns exemples es mostren en (5).
 - (5) a. Ell escalava.
 - b. Ell està escalant.

- **Processos duratius acabats:** Esdeveniments que impliquen una duració amb un final definit en el qual un estat canvia. En l'exemple (6), l'estat de *estar escalant* canvia en el punt de culminació de *estar al cim*.
 - (6) a. Ell va escalar fins al cim.
 - b. Ell va escalar fins al cim en dues hores.

- **Acabaments no duratius:** Esdeveniments puntuals o instantanis que no impliquen una duració, i que tenen lloc en un moment precís en el qual l'estat canvia. Alguns exemples es mostren en (7).
 - (7) a. Ell ha arribat al cim.
 - b. Ha escalat 200 m.

Per a entendre aquesta classificació, els conceptes de **dinamisme**, **acabament**, i **duració** són fonamentals (Smith, 1991). Com mostra la taula 2.2, el dinamisme separa els estats, que descriuen propietats estàtiques, de la resta de categories, que descriuen situacions dinàmiques (esdeveniments); l'acabament indica si un esdeveniment implica un canvi d'estat; i la duració defineix si un esdeveniment ocupa un interval en el temps o és puntual (*i. e.*, ocupa només un instant).

| Categoria aspectual | Dinamisme | Acabament | Duració | Exemple |
|----------------------|-----------|-----------|---------|--------------------|
| Estat | no | no | sí | saber, tenir |
| Procés/Activitat | sí | no | sí | caminar, pintar |
| Acabament duratiu | sí | sí | sí | construir, trencar |
| Acabament no duratiu | sí | sí | no | advertir, guanyar |

Taula 2.2: Propietats de les categories aspectuals

L'aspecte gramatical i els adverbis temporals (per exemple, *ahir*, *durant dues hores*) són recursos lingüístics utilitzats per a canviar entre els aspectes lèxics. Per exemple, la forma progressiva expressa els processos (p. e.: *John està menjant*), mentre que la no-progressiva (perfectiva) generalment

expressa la culminació de les accions (per exemple, *John ha trobat la seua clau*).

Moens i Steedman (1988) van presentar una ontologia d'esdeveniments influent seguint la classificació de Vendler. Hi presentaren una estructura tripartida d'esdeveniments basada en el concepte de *contingència* (vegeu la figura 2.1).

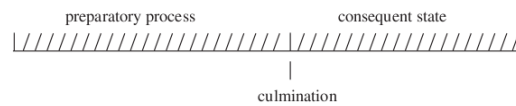


Figura 2.1: Estructura (nucli) de l'esdeveniment basat en la contingència

L'ontologia inclou un model computacional per al càlcul dels aspectes dels esdeveniments. Les propietats de les categories aspectuals i els canvis d'aspecte lèxic entre ells van ser modelats en una xarxa de transició (vegeu la figura 2.2).

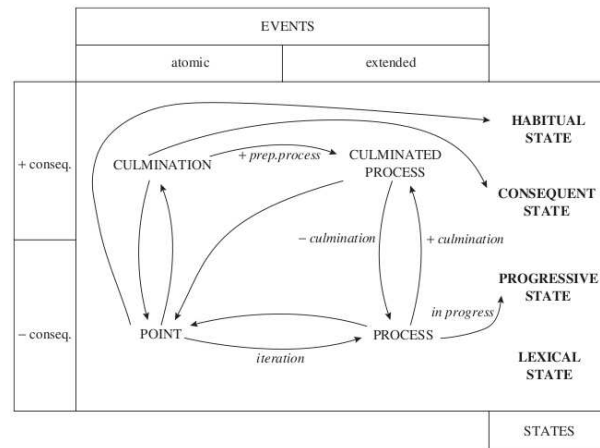


Figura 2.2: Xarxa de transaccions per al càlcul aspectual

Com mostra la figura, el model afegeix una categoria aspectual a la classificació de Vendler per a esdeveniments puntuals on les conseqüències no són rellevants: el punt (EN: *point*) (per exemple, *tossir*).

Modalitat

D'acord amb Cristall (2003), el mode ('modalitat') es refereix a un conjunt de restriccions sintàctiques i semàntiques del verb (per exemple, indicatiu, subjuntiu, imperatiu, interrogatiu), i representa l'actitud per part del parlant cap al contingut objectiu de l'emissió (per exemple, incertesa, definició, possibilitat, etc.). El mode s'assenyala mitjançant formes alternatives d'inflexió verbal, o mitjançant l'ús d'auxiliars.

A més del temps verbal i l'aspecte, en anglès (i en altres idiomes), la modalitat també participa en l'expressió lingüística de temps. Per exemple, l'expressió dels esdeveniments futurs està estretament relacionada amb la interacció entre la tensió i la modalitat (Lyons, 1968).

Raonament temporal

Raonar sobre el temps implica la representació temporal dels esdeveniments i les seues relacions temporals. El desenvolupament de models computacionals eficients per al raonament sobre el temps ha estat una àrea central de la recerca en intel·ligència artificial des dels anys 60.

L'*Interval Algebra* d'Allen (1983) és una de les obres més influents en aquesta línia. En el seu model, els esdeveniments són representats pels intervals de temps. Allen es defineix 13 possibles relacions entre dos intervals: **igual** (=), **abans** (<), **després** (>), **immediatament abans** (*m*), **immediatament després** (*mi*), **superposa** (*o*), **superposat per** (*oi*), **comença** (*s*), **començat per** (*si*), **acaba** (*f*), **acabat per** (*fi*), **durant** (*d*), **conté** (*di*); vegeu la figura 2.3.


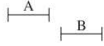





| | |
|---|--------------------------------------|
|  | A is EQUAL to B B is EQUAL to A |
|  | A is BEFORE B B is AFTER A |
|  | A MEETS B B is MET by A |
|  | A OVERLAPS B B is OVERLAPPED by A |
|  | A STARTS B B is STARTED by A |
|  | A FINISHES B B is FINISHED by A |
|  | A DURING B B CONTAINS A |

Figura 2.3: Les 13 relacions temporals d'Allen

CAPÍTOL 2. ESTAT DE LA QÜESTIÓ

La figura 2.4 mostra la representació en intervals dels esdeveniments descrits en (8).

- (8) Mary hit the ball while playing football.
The balloon exploded.



Figura 2.4: Relacions temporals

En aquest exemple, *hitting* ocorre **durant** *playing*. Amés, *hitting* ocorre **immediatament-abans** que *exploding*, el qual **acaba** *playing*.

En el model d'Allen, les relacions temporals entre els intervals es mantenen en una xarxa, on els nodes representen els intervals individuals i els arcs representen la relació entre ells com mostra la figura 2.5.

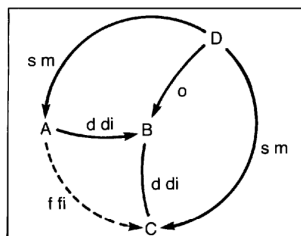


Figura 2.5: Xarxa de relacions temporals

Se suposa que la xarxa sempre manté la informació completa sobre com estan relacionats els intervals. Quan s'introdueix un nou interval es creen noves relacions, i totes les conseqüències es calculen pel tancament transitiu de les relacions temporals i la consistència de la xarxa es calcula mitjançant un algorisme de propagació de restriccions. Aquest consisteix a afegir noves relacions consistents entre els intervals que no estan relacionats seguint una taula de transitivitat (vegeu la figura 2.6).

| B r2 C | < | > | d |
|---------------|------------|------------|-------------------|
| A r1 B | | | |
| "before" < | < | no info | < o m d s |
| "after" > | no info | > | > oi mi d f |
| "during" d | < | > | d |

Figura 2.6: Part de la taula de transitivitat d'Allen

L'exclusió dels instants i la complexitat computacional de la proposta d'Allen van ser criticades per Vilain et al. (1982, 1990). Ells van presentar l'Algebra Puntual, que introdueix instants (punts) com a unitat bàsica en la proposta d'Allen. Un interval pot ser definit com un parell d'instants (punt inicial i punt final), com es mostra en la figura 2.7. Això tradueix la relació entre dos intervals (X i Y) en sis relacions entre els quatre punts resultants (x_1 , x_2 , y_1 i y_2). Atès que el punt inicial sempre precedeix ($<$) el punt final, hi ha només quatre relacions de punt (de R1 a R4), també anomenades *convex relations*.

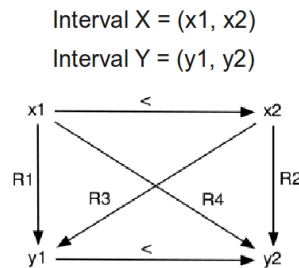


Figura 2.7: Descomposició de una relació entre intervals en relacions entre punts

Després de la traducció dels intervals en punts, el nombre de relacions augmenta d'una relació d'interval fins a quatre relacions de punt. Però l'avantatge d'aquest enfocament és que, si bé hi ha tretze possibles relacions entre els intervals, només hi ha tres possibles relacions entre els punts (és a dir, $=$, $<$, $>$). Per tant, la taula de la transitivitat és més simple i llavors

l'algorisme és més eficient.

Altres notables millores d'aquests models en termes d'eficiència són Freska (1992) (*Conceptual Neighborhood*) i Gerevini et al. (Gerevini, Schubert, & Schaeffer, 1993) (*Timegraph*). No obstant això, l'eficiència i l'aplicació de models de raonament temporal estan fora de l'abast d'aquesta tesi. L'element central d'aquesta tesi és el conjunt de les relacions temporals que dos esdeveniments poden presentar, siguen duratius o instantanis.

Estructura temporal del discurs

El context d'un discurs és necessari per a assolir la comprensió del llenguatge, incloent-hi l'ordenació temporal dels esdeveniments. Per exemple, en (9), seguint el sistema de Reichenbach, l'esdeveniment *guanyat* seria anterior al moment de la conversa, mentre que en realitat la frase només diu que és anterior a l'esdeveniment *pensarà*.

(9) Ell pensarà que ella ha guanyat.

Treballs remarcables sobre l'estructura del discurs són el de Dowty (1986) sobre el principi d'interpretació del discurs (TDIP); Webber (1988) amb l'estudi de la anàfora del discurs; Hwang and Schubert (1992) amb els *Tense Trees*; Kamp and Reyle (1993) amb el *Reference Point* (Rpt); els arbres aspectuals dinàmics (DATs) (ter Meulen, 1995); i els gràfics conceptuals temporals (Moulin, 1997). Les influències pragmàtiques del discurs, com la causalitat, també han sigut considerades. Lascarides i Asher (1993) van ressaltar que el sentit comú i el coneixement sobre el món real són necessaris per a la correcta interpretació de les relacions temporals. En els exemples (10a) i (10b), la causalitat hi té un paper important en la inferència de l'ordre dels esdeveniments. Cada exemple presenta una ordenació textual diferent, però el lector pot inferir que l'esdeveniment *empentar* precedeix el de *caure* en ambdós casos.

- (10) a. John empentà a Marc. Marc caigué.
b. Marc caigué. John l'empentà.

En aquesta subsecció es posa en relleu els estudis de la lingüística teòrica que fan front a l'expressió lingüística de les relacions temporals complexes en els nivells de discurs i pragmàtica. Els estudis assenyalen que és necessari considerar aquests nivells per a arribar a un perfecte rendiment en el processament de la informació temporal. No obstant això, aquesta tesi se centra en la semàntica, i per tant, el discurs i les influències pragmàtiques estan fora de l'abast de la nostra proposta.

2.2 Metodologia actual i anotació temporal

Els models formals descrits anteriorment en la secció 2.1 van ser criticats perquè no es van analitzar i avaluar amb dades lingüístiques reals. A mitjans dels anys 90, la potència de càlcul i la disponibilitat de textos digitalitzats van experimentar un creixement important. Això va produir un canvi epistemològic en el processament de la informació temporal i, en general, del PLN va passar d'una estratègia racionalista basada en les teories formals d'anàlisi del llenguatge a una estratègia empírica basada en l'anàlisi de l'ús del llenguatge real (és a dir, corpus textual) (Manning & Schütze, 1999).

L'avantatge de la lingüística empírica o de corpus és que posa de manifest tant la varietat de formes d'expressió com la seua freqüència en una mostra real de la llengua (McEnery & Wilson, 1996).

Els corpus guien el desenvolupament de models computacionals i proporciona un marc d'avaluació objectiva. A més, els corpus poden ser compartits, discutits, re-usats i corregits. No obstant això, la lingüística de corpus també presenta un desavantatge: l'escassetat de dades. Es necessita una gran quantitat de dades de text per a fer generalitzacions i extraure conclusions fiables sobre el llenguatge. D'altra banda, donat el caràcter creatiu de la llengua, és molt difícil construir un corpus amb una àmplia cobertura de les expressions lingüístiques.

Utilitzant només un corpus es pot fer processament lingüístic superficial. Però, l'anotació d'aquest corpus amb informació lingüística (per exemple, sintàctica, semàntica, temporal, etc.) permet l'anàlisi dels coneixements lingüístics que contenen. Per tal d'anotar corpus, es necessita un esquema d'anotació. Un esquema d'anotació defineix quines instàncies d'una classe particular de fenòmens han de ser anotats en els corpus i com aquestes han de ser anotades. L'esquema també ha d'incloure una documentació detallada (guia) sobre com usar-lo correctament.

Centrant-nos en la informació temporal, els treballs formals sobre el temps i el llenguatge s'utilitzaren com a base per a desenvolupar esquemes d'anotació temporal. Els esquemes d'anotació temporal més destacats són: MUC-TIMEX (Grishman & Sundheim, 1996), TIDES (Ferro, Mani, Sundheim, & Wilson, 2000), STAG (Setzer & Gaizauskas, 2000), i TimeML (Pustejovsky, Castaño, Ingria, Saurí, Gaizauskas, Setzer, & Katz, 2003). Tots segueixen un format d'anotació basat en SGML/XML. Aquests esquemes es descriuen a continuació en ordre cronològic, destacant les novetats que introdueix cada un sobre el seu predecessor.

1995/97 – MUC-TIMEX

L'exemple més primerenc d'anotació d'informació temporal es remunta a 1995, quan la tasca d'entitats anomenades inclou el reconeixement d'expressions temporals absolutes en la 6a Conferència de DARPA (MUC-6) (Grishman & Sundheim, 1996). A diferència dels treballs teòrics, que se centaven en les propietats temporals dels esdeveniments, aquest esquema temporal se centra primer en l'anotació dels adverbis temporals i sintagmes temporals que representen dates concretes (per exemple, setembre de 1999) i hores (per exemple, 17:00). Al MUC, als participants se'ls va demanar que construïren sistemes que marcaren aquestes expressions amb l'etiqueta **TIMEX**⁴ i que n'indicaren el tipus (data o hora) en un corpus textual. Un exemple es mostra en (11).

(11) The rocket was launched on
<TIMEX type="DATE">October 1st, 1999</TIMEX>

Per tal de mesurar l'eficàcia dels enfocaments participants, els seus resultats van ser comparats amb un corpus anotat manualment (*gold standard*).

Posteriorment, al 1997 al MUC-7 (Gaizauskas & Wilks, 1998), la tasca de reconeixement de TIMEX inclogué les expressions temporals relatives⁵ (p. e.: *ahir, fa dos anys*).

Aquest esquema d'anotació inicial era molt simple, limitat a la identificació d'expressions temporals en els textos.

2000 – TIDES

Amb el suport de DARPA i ACE, l'esquema Detecció, Extracció i Resum d'Informació Translingüe TIDES (Ferro et al., 2000; Wilson et al., 2001; Ferro et al., 2005) va sorgir amb el propòsit de l'anotació multilingüe d'expressions temporals. Això va substituir l'etiqueta XML TIMEX per la TIMEX2. Els principals avantatges que introdueix als atributs de l'anterior etiqueta TIMEX són:

- A més de dates i hores, es consideren les **duracions** (DURATION type) (p. e.: *durant dos anys*) i **repeticions** (SET type) (p. e.: *mensualment*).
- Les dates absolutes (p. e.: *October 1st, 1999*), relatives (p. e.: *yesterday*), períodes (p. e.: *two years*), i repeticions temporals (p. e.:

⁴El mot TIMEX és un acrònim de *expressió del temps*.

⁵Relatives al temps la creació de documents (DCT).

weekly) han de ser **normalitzades** en un nou atribut VAL⁶, seguint l'estàndard ISO 8601.

- S'afegeixen **atributs addicionals** que capturen la semàntica de les *timex*:
 - *MOD*: captura els modificadors temporals.
 - *ANCHOR_VAL*: Conté una data/hora de referència relacionada.
 - *ANCHOR_DIR*: Captura la direcció temporal entre VAL i ANCHOR_VAL.
 - *SET*: Identifica expressions que denoten repetició.

Es mostra un exemple en (12):

- (12) A rocket was launched <TIMEX2 VAL="1999-10">in October, 1999</TIMEX2>. Three rockets were launched <TIMEX2 VAL="P1Y" ANCHOR_VAL="2000">during the next year</TIMEX2>. She visits him <TIMEX2 SET="YES" VAL="XXXX-XX-XX">daily</TIMEX2>. She knows him <TIMEX2 VAL="P1Y" MOD="LESS_THAN">for less than a year</TIMEX2>.

Aquest esquema d'anotació va ser emprat en el fòrum d'avaluació TERN (2004) i en l'EVALITA'07⁷. TIDES va ser l'esquema més important (Negri & Marseglia, 2004; Saquete et al., 2006) fins l'adopció del TimeML com a estàndard.

El problema d'aquest esquema és la seua limitació a l'anotació d'expressions temporals. Això exclou els esdeveniments i les relacions temporals que són entitats importants per al raonament temporal.

2000/01 – Sheffield Temporal Annotation Guidelines (STAG)

Aquest esquema va ser motivat per les obres contemporànies en el corpus basat en processament de la informació temporal.

Al taller de processament d'informació temporal i espacial de l'ACL-2001 es van presentar alguns enfocaments influents per a l'anglès (Filatova & Hovy, 2001; Katz & Arosio, 2001) i l'alemany (Schilder & Habel, 2001). Aquests suggerien l'extensió de l'anotació de la informació temporal de només TIMEX a altres entitats temporals, com els esdeveniments i les relacions temporals, per a construir enfocaments amb capacitats de raonament temporal.

⁶L'atribut *type* desapareix.

⁷<http://www.evalita.it/2007/tasks/tern>

L'STAG resol aquest problema proporcionant una visió més completa de l'esquema d'anotació de la informació temporal. Va ser desenvolupat per Setzer i Gaizauskas (2000, 2001) amb l'objectiu d'identificar els esdeveniments en les notícies, així com la localització temporal i l'ordre relatiu d'aquestes. L'STAG es compon de les següents etiquetes: TIMEX (heredada del MUC i de TIDES), SIGNAL i EVENT. Es mostra un exemple en (13).

(13) A small single-engine plane <event eid="9" class="OCCURRENCE" tense="past" relatedToTime="5" timeRelType="included" signal="9"> crashed </event> into the Atlantic Ocean about eight miles off New Jersey <signal sid="9">on</signal> <timex tid="5" type="DATE" calDate="12031997">Wednesday</timex> .

Les innovacions introduïdes per l'STAG van ser:

- L'anotació de l'expressió lingüística dels **esdeveniments** (*i. e.*, coses que ocorren) amb l'etiqueta EVENT, en la qual s'especificaven com a atributs les seues relacions i altres propietats com la classe. Es van definir quatre classes d'esdeveniments: **occurrence**, **perception**, **reporting** i **aspectual**. Els esdeveniments en forma d'estats definits en els treballs de la lingüística formal (Vendler, 1967) no es van incloure en l'esquema a causa de la seua complexitat.
- L'anotació d'**assenyaladors temporals**. Aquests elements assenyalen la relació entre les entitats temporals (timex-event, timex-timex, o event-event). Es denoten per preposicions (p. e.: *on*, *after*, *during*) i conjuncions (p. e.: *while*, *when*). Per exemple, en (13) *on* indica que l'esdeveniment *crashing* es relaciona amb *Wednesday*.
- L'etiqueta EVENT inclou les **relacions temporals**. Aquests atributs indiquen les relacions d'un esdeveniment amb altres esdeveniments o expressions temporals. A més, els tipus de les relacions estan basats en les 13 relacions temporals definides per Allen (1983). Per exemple en (13), *crash* és un esdeveniment *inclòs* (*during*) en l'interval de temps representat per *Wednesday*.

Com s'ha indicat, aquest esquema no inclou els esdeveniments d'estat i, a més, les relacions temporals s'inclouen en l'etiqueta EVENT, cosa que va produir alguns problemes en l'aplicació de l'esquema als textos reals que contenen relacions temporals complexes.

2002/03 – Time Markup Language (TimeML)

TimeML⁸ (Pustejovsky et al., 2002, 2003) es va desenvolupar sota el patrocini d'ARDA com l'evolució natural d'STAG. TimeML és un llenguatge d'especificació ric i complet per a l'anotació de esdeveniments, expressions temporals i les seues relacions en el text en llenguatge natural.

L'exemple (14) mostra una frase anotada amb TimeML. Aquesta frase inclou una relació temporal (TLINK) entre l'esdeveniment (EVENT) *came* i l'expressió temporal (TIMEX3) *Monday*, en la qual l'assenyalador (SIGNAL) *on* participa.

```
(14) John <EVENT eid="e1" class="occurrence">came</EVENT>
      <SIGNAL sid="s1">on</SIGNAL>
      <TIMEX3 tid="t1" type="dat" value="2010-06-2"> Monday </TIMEX3>

      <TLINK reltype="includes" relatedToEvent="e1"
      timeID="t1" signal="s1">
```

TimeML combina i estén les característiques d'ambdós esquemes anteriors, la qual cosa el converteix en un esquema d'anotació més potent. En l'actualitat, TimeML ha estat adoptat com a estàndard per la majoria dels investigadors per a anotar la informació temporal, per la seua integritat i les millores que afegeix als esquemes anteriors. A més, ha estat proposat com a estàndard ISO⁹ (ISO-TimeML, 2007).

El TimeML captura l'expressió lingüística de les següents entitats temporals:

- **Expressions temporals (*timex*).** Una expressió temporal o de temps (*timex*) és una representació lingüística d'un punt o període en el temps. Normalment la denota un substantiu, adjectiu, adverb, o un sintagma nominal, adjectival, adverbial o verbal. L'exemple (15) mostra alguns TIMEX subratllats.

- (15) a. John came on Monday and he will leave on October 9th, 2012.
 b. The show starts at 8 p.m.
 c. She will be here for two weeks.
 d. I visit my parents monthly.

En (15a), *Monday* i *October 9th, 2012* són una *timex* relativa i absoluta respectivament que descriuen un punt específic en el temps. En

⁸<http://timeml.org>

⁹http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf

(15b), *8 p.m.* és una expressió temporal amb granularitat d'hores. L'exemple (15c) inclou un sintagma nominal *two weeks* el qual representa un període. Finalment, l'exemple (15d) mostra una expressió temporal recurrent.

D'acord amb TimeML, TIMEX es poden classificar en quatre tipus:

- *Date (data)*. Referències temporals amb una granularitat de dies, setmanes, i majors, per exemple (15a).
- *Time (hora)*. Referències temporals amb una granularitat d'hores, minuts, i menors, per exemple (15b).
- *Duration (duració)*. Períodes de temps, com (15c).
- *Set (repetició)*. Patrons que es repeteixen en el temps com (15d).

A més, per a facilitar la interpretació de càlcul d'expressions temporals, es va afegir un atribut (*value*) a TIMEX per a capturar la seua normalització ISO 8601. Alguns exemples es mostren en (16).

- (16) a. October 2012: 2012-10
b. Two weeks: P2W
c. Monthly: XXXX-XX
d. Tomorrow at 1 p.m.: depends on the DCT

L'exemple (16a) consisteix en una data explícita (tipus DATE) que especifica l'any i el mes. L'exemple (16b) és una duració. Les duracions es normalitzen emprant tres elements: *P* de període, un nombre, i una abreviatura d'una unitat temporal (p. e.: week → W). En l'exemple s'obté *P2W* per a *period of 2 weeks*. Si la duració conté més d'una unitat temporal (p. e.: two weeks, one day, and one hour) la normalització es concatena *P2W1DT1H*. Si la duració no defineix un nombre concret (p. e.: some weeks) aquest es deixa com a X. L'exemple (16c) és una *timex* recurrent (tipus SET). Els *sets* es normalitzen com a patrons ISO 8601 deixant com a X els valors en els quals l'expressió pot variar. En l'exemple, el resultat és XXXX-XX, que significa que l'any (XXXX) i el mes (XX) poden tenir qualsevol valor (monthly = tots els mesos de tots els anys). Finalment, en (16d), es mostra una hora relativa. Per a normalitzar-la, cal tenir en compte una referència temporal de la data en la qual es va crear el text on apareix (DCT – date creation time). Si la DCT és per exemple 2010-06-30, el valor ISO 8601 seria 2010-07-01T13:00.

La definició TimeML de les *timex* es basa principalment en TIMEX2, però a causa de les diferències, introdueix una nova etiqueta: TIMEX3. L'extensió de la *timex* ha de correspondre a les següents categories: sintagma nominal (*Yesterday*), sintagma adjectival (*3 days*) o adverbial (*fairly recently*). Per tant, qualsevol preposició precedent no està inclosa com a part de l'element. A més, alguns modificadors adverbials (per exemple, *ago*, *ever*) es consideren part de la TIMEX, però no els modificadors que expressen un esdeveniment (*after he came back*).

- **Esdeveniments.** Un esdeveniment es defineix com l'expressió lingüística d'alguna cosa que passa, o alguna propietat dinàmica que obté un valor durant un període de temps (estat). Els esdeveniments poden ser puntuals o durar un període de temps. En el llenguatge natural, els esdeveniments s'expressen generalment mitjançant verbs finits (17a) i no-finites (17b). El verb és la part més natural d'expressió per a expressar esdeveniments (Levin & Hovav, 2005). No obstant això, els esdeveniments poden ser expressats també pels noms. Els substantius poden realitzar esdeveniments de tres formes diferents (Gross & Kiefer, 1995): nominalitzacions de verbs (17c), esdeveniments nominals (17d), i noms que denoten esdeveniments a causa d'un *type-coercion* (Pustejovsky, 1995) induït pels verbs (17e). A més, els esdeveniments (especialment els estats), poden ser representats per adjectius (17f), clàusules predicatives (17g) i locucions prepositives (17h).

- (17) a. John came on Monday.
 b. He called the teacher to thank him.
 c. U.S. will delay the attack.
 d. This will end the war.
 e. The beer lasted only two seconds. We came back after the beer.
 f. The volcano, dormant for ten years, is now erupting.
 g. The party will be ready on time.
 h. All 75 people on board the Airbus survived.

D'acord amb TimeML, hi ha set classes d'esdeveniments:

- *Reporting*. Acció de narrar un esdeveniment (p. e.: *say*, *report*).
- *Perception*. Percepció física d'un altre esdeveniment (p. e.: *see*, *hear*).
- *Aspectual*. Predicació aspectual d'un altre esdeveniment (p. e.: *start*, *continue*).

- *I_Action*. Acció intensional (p. e.: *try, attempt*).
- *I_State*. Estat intensional (p. e.: *feel, hope*).
- *State*. Circumstància en la qual alguna propietat dinàmica és certa (p. e.: *rainy, bad*).
- *Occurrence*. Esdeveniments que descriuen coses que passen (p. e.: *came, war*).

L'etiqueta **EVENT** s'hereta de l'**STAG** i s'estén amb les classes **state**, **i.action** and **i.state**. Els atributs de l'**STAG** que s'empraven per a capturar les relacions temporals són eliminats, ja que aquestes es representen de manera separada en altres etiquetes per a millorar el tractament de les relacions complexes.

- **Relacions temporals**. En el llenguatge natural, hi ha relacions entre un esdeveniment i una *timex*, entre dos esdeveniments, i entre dues *timex*, en les quals els assenyaladors temporals (**SIGNAL**) poden participar. TimeML defineix l'etiqueta **TLINK** per a les relacions temporals. A més, TimeML captura altres tipus de relacions: la subordinació (**SLINK**) i l'aspectual (**ALINK**).

Un **TLINK** estableix un vincle entre dues entitats per a explicitar-ne la seua relació temporal. Prenent com a base l'**STAG**, en TimeML, els tipus de relació es basen també en les relacions d'Allen:

- *simultani*: dues entitats que es produeixen al mateix temps o estan temporalment indistingibles en el context – vegeu l'exemple (18). En l'exemple, l'esdeveniment *teaching* i el període *20 minutes* són simultanis.
(18) John taught for 20 minutes on Monday.
- *abans*: una entitat ocorre abans que l'altra – vegeu l'exemple (19). En l'exemple, *murder* ocorren abans que *arresting*.
(19) The police looked into the murders of 14 women.
In six of the cases suspects have already been arrested.
- *després*: La categoria inversa de l'anterior.
- *immediatament abans (meets)*: El final d'un esdeveniment coincideix en el temps amb el començament de l'altre – vegeu l'exemple (20). En l'exemple, el xoc és immediatament abans de les morts (ambdós esdeveniments s'encadenen en el temps).

2.2. ESQUEMES D'ANOTACIÓ TEMPORAL

(20) All passengers died when the plane crashed into the mountain.

- *immediatament després (meet-by)*: La categoria inversa de l'anterior.
- *inclou parcialment (overlap)*: L'extensió de temps d'una entitat n'inclou parcialment una altra – vegeu (21).

(21) After a two-day journey, John arrived in Boston last Thursday.

En l'exemple, *journey* inclou parcialment *last Thursday*. Per exemple, si el viatge va acabar durant la vesprada, comença abans que *last Thursday* (dos dies abans) i acaba abans del final de *last Thursday* (la vesprada).

- *inclòs parcialment (overlapped-by)*: La categoria inversa de l'anterior.
 - *durant*: Una entitat està completament compresa entre els límits temporals d'una altra que dura més temps.
- (22) a. Two cars crashed during the race.
b. John taught for 20 minutes on *Monday*.
- *conté*: La categoria inversa de l'anterior.
 - *comença*: Com la relació *durant*, però el límit esquerre coincideix – vegeu l'exemple (23). En l'exemple, *6:00 pm* comença l'estat *in the gym*.

(23) John was in the gym between 6:00 p.m and 7:00 p.m.

- *començat-per*: La categoria inversa de l'anterior.
 - *acaba*: Com la relació *durant*, però el límit dret coincideix – vegeu l'exemple (24). En l'exemple, *7:00 pm* acaba l'estat *in the gym*.
- (24) John was in the gym between 6:00 p.m and 7:00 p.m.

- *acabat-per*: La categoria inversa de l'anterior.
- *identitat*: S'utilitza per a marcar esdeveniments que corresponen al mateix esdeveniment – vegeu l'exemple (25). En l'exemple, ambdós *drove* i *drive* són el mateix esdeveniment. Per tant, estan relacionats per un TLINK d'identitat.

(25) John drove to Boston.

During his drive he ate a donut.

Un **SLINK** defineix una relació entre dos esdeveniments en una subordinació sintàctica – vegeu (26). Els SLINKs són d'algun dels següents tipus: *modal*, *factive*, *counter-factive*, *evidential*, *negative-evidential*, o *conditional*.

```
(26) John <EVENT eid="e1">said</EVENT>  
      he <EVENT eid="e2">bought</EVENT> some wine.  
  
      <SLINK eventInstanceID="e1" subordinatedEventInstance="e2"  
      relType="EVIDENTIAL"/>
```

Un **ALINK** representa la relació entre un esdeveniment aspectual i l'esdeveniment del seu argument – vegeu (27). Els ALINKs són d'algun dels tipus següents: *initiation*, *culmination*, *termination*, *continuation*, or *reinitiation*.

```
(27) John <EVENT eid="e1">started</EVENT> to  
      <EVENT eid="e2">read</EVENT>.  
  
      <ALINK eventInstanceID="e1" relatedToEventInstance="e2"  
      relType="INITIATE"/>
```

L'esquema TimeML va ser originalment desenvolupat per a l'anglès i va ser llançat en conjunt amb un corpus anotat: TimeBank (Pustejovsky et al., 2003). Però hi ha hagut moltes iniciatives per a estendre l'esquema a altres llengües com el francès (Bittar, 2009), l'italià (Caselli, 2009), el coreà (Im et al., 2009), el romanès (Forascu, 2008), el portuguès (Costa & Branco, 2010), turc (Seker & Diri, 2010), i també el castellà (Saurí, 2010) i el xinès (Xue & Zhou, 2010). Idiomes que van ser inclosos en l'exercici d'avaluació internacional TempEval-2 (Verhagen et al., 2010).

Les etiquetes XML i els atributs dels sistemes de revisió prèviament apareixen resumits i comparats en la taula 2.3.

2.2. ESQUEMES D'ANOTACIÓ TEMPORAL

| | MUC-6&7 1995/1997 | TIDES 2000 | STAG 2000/2001 | TimeML 2002/2003 |
|------------------------------|---|--|--|--|
| TIMEX | TIMEX <u>attributes</u> type | TIMEX2 <u>attributes</u> - val set mod anchor_dir anchor_val | TIMEX <u>attributes</u> type calDate - - - tid | TIMEX3 <u>attributes</u> type value - mod - tid temporalFunction anchorTime functionInDoc |
| EVENT | | | EVENT <u>attributes</u> eid class tense/aspect <i>relatedToEvent</i> <i>eventRelType</i> <i>relatedToTime</i> <i>timeRelType</i> <i>signal</i> | EVENT <u>attributes</u> eid class tense/aspect |
| SIGNAL | | | SIGNAL <u>attributes</u> sid | SIGNAL <u>attributes</u> sid |
| TEMPORAL RELATION | | | (<i>included in EVENT</i>) | [T,A,S]LINK <u>attributes</u> lid eid or tid relatedTo eid or tid sid relType |

Taula 2.3: Comparativa del esquemes d'anotació temporal

Tingueu en compte que els atributs TimeML que apareixen descrits en la taula s'han simplificat. Per a més detalls, vegeu <http://timeml.org>.

En l'actualitat, TimeML és l'esquema d'anotació estàndard per al processament de la informació temporal, ja que és l'esquema més complet i millora les característiques dels seus predecessors. En el següent apartat es descriuen els mètodes computacionals actuals d'anotació TimeML automàtica.

2.3 Aproximacions computacionals actuals

Avui dia, el processament de la informació temporal es compon de: donat un text d'entrada en llenguatge natural, anotar-lo automàticament seguint d'un esquema d'anotació temporal. La majoria de les propostes computacionals recents s'han desenvolupat per a l'anglès i segueixen l'esquema d'anotació TimeML, que és l'esquema estàndard actual i un futur estàndard ISO.

Aproximacions per a l'anglès

En ordre cronològic, els enfocaments més rellevants d'anotació automàtica TimeML sobre text en anglès són els següents.

Abans del TempEval

Els primers sistemes a realitzar aquesta tasca es van avaluar i analitzar de forma aïllada, perquè en aquell moment no havia hagut cap marc d'avaluació internacional compartida, com ara TempEval (Verhagen et al., 2007, 2010). En aquesta etapa, els sistemes més rellevants són els següents.

TTK o TARSQI Tool Kit (Verhagen et al., 2005) és un sistema complet d'anotació TimeML. Implementa cinc mòduls principalment basats en regles: GUTime, EVITA, GUTenLINK, Slinket i SputLink. GUTime és el mòdul de reconeixement de TIMEX3. Aquest mòdul es basa en el Tempex tagger (Mani & Wilson, 2000). Evita és el mòdul encarregat del reconeixement d'esdeveniments. Les seues regles es basen en: categories gramaticals, lematització i fragmentació obtinguts mitjançant l'Alembic Workbench; la informació lèxica i l'anàlisi contextual; i WordNet¹⁰ en combinació amb la desambiguació apresada amb un mètode bayesià per a la identificació d'esdeveniments nominals. GUTenLINK és un etiquetador de TLINKs basat en la informació sintàctica i lèxica. Slinket identifica les relacions de subordinació entre parells d'esdeveniments que estan representats per l'etiqueta SLINK en TimeML. Finalment, SputLink és un component de tancament temporal que estén les relacions temporals conegudes amb noves relacions.

Boguraev i Ando (2005) presentaren un enfocament basat en aprenentatge automàtic sobre característiques morfosintàctiques. Van desenvolupar un anotador automàtic de TimeML utilitzant tècniques d'aprenentatge automàtic (ML) sobre el corpus TimeBank. L'anotació de TIMEX3 i SIGNAL es realitza mitjançant gramàtiques d'estats finits en cascada. L'anotació d'esdeveniments i TLINKs es du a terme mitjançant un classificador de minimització de riscos robust (MRR).

¹⁰<http://wordnet.princeton.edu>

Mani et al. (2006) presentà un sistema de categorització de les relacions temporals, que compara diferents tècniques de ML: màquines de suport vectorial (SVM), màxima entropia i Naive Bayes, utilitzant característiques morfosintàctiques. En aquesta línia, Lapata i Lascarides (2006) van presentar un estudi de diferents estratègies comparant models probabilístics conjuntius i disjuntius.

STEP (Bethard & Martin, 2006) és un sistema de reconeixement i classificació d'esdeveniments. Aquest mètode utilitza un ampli conjunt de característiques textuais, morfològiques, de dependència sintàctica i WordNet (hiperonímia) per a construir un model de SVM. En aquesta línia, March i Baldwin (2008) presentaren un sistema de reconeixement d'esdeveniments mitjançant un classificador multiclasse (també SVM). Les principals característiques utilitzades per a entrenar el classificador són paraules i categories gramaticals en una finestra de context, eliminació de paraules comunes i generalització de característiques a través d'agrupació de les paraules (els nombres, les entitats esmentades, etc.)

En els enfocaments revisats, la comparació equitativa entre els resultats no va ser possible perquè els autors van construir el seu propi marc d'avaluació. Per aquesta raó, es van organitzar els fòrums internacionals d'avaluació, per a mesurar i comparar el rendiment en el processament de la informació temporal (basada en TimeML). Hi ha hagut dos exercicis d'avaluació importants: TempEval-1 (2007), i TempEval 2 (2010).

TempEval-1 (Verhagen et al., 2007)

Aquesta avaluació es va centrar només en la categorització automàtica de tres tipus de relacions temporals TimeML: intraoracionals TIMEX esdeveniments, esdeveniment-DCT, i entre esdeveniments de diferents oracions. Aquesta primera avaluació es va limitar a aquestes relacions i el processament de TIMEX i esdeveniments va ser exclòs a causa de la complexitat i les dimensions de dur a terme totes les tasques. A més, només es va avaluar l'idioma anglès pel fet que només hi havia dades TimeML en anglès disponibles en aquell moment. Els sistemes que van participar en TempEval-1 foren els següents.

WVALI (Puscasu, 2007) i XRCE-T (Hagège & Tannier, 2007) són sistemes basats en regles que utilitzen la informació morfosintàctica. En les seues descripcions, inclouen les classes de les *timex* i dels esdeveniments com a trets semàntics.

CU-TMP (Bethard & Martin, 2007) és un sistema basat en màquines de suport vectorial (SVM), que utilitza les característiques morfosintàctiques.

USFD (Hepple, Setzer, & Gaizauskas, 2007) és també un sistema d'a-

prementatge automàtic, que utilitza les característiques morfosintàctiques, basat en dues tècniques de ML (SVM i Naive Bayes).

NAIST (Cheng, Asahara, & Matsumoto, 2007) difereix dels enfocaments anteriors, ja que utilitza una tècnica de ML que combina els models ocults de Markov i SVM.

Finalment, LCC-TE (Min, Srikanth, & Fowler, 2007) és un sistema híbrid que combina les regles i diferents tècniques ML (SVM, Naive-Bayes, l'entropia màxima, i arbres de decisió). També utilitza la semàntica lèxica i tècniques de desambiguació del sentit de les paraules.

TempEval-2 (Verhagen et al., 2010)

Aquesta avaluació inclou *timex*, esdeveniments i relacions esdeveniment-*timex*, DCT-esdeveniment, esdeveniment principal, i les relacions de subordinació d'esdeveniments. Aquest és actualment l'exercici d'avaluació més important en el processament de la informació temporal. A més, TempEval-2 suposà un gran esforç d'anotació: a més d'anglès, es van anotar corpus per a l'espanyol, l'italià, el francès, el coreà i el xinès. Els següents sistemes van participar en TempEval-2 en anglès¹¹.

HeidelTime (Strötgen & Gertz, 2010) només es va dirigir a la tasca de processament de *timex* i va obtenir la millor puntuació. Aquest sistema es basa en regles fetes a mà sobre característiques morfosintàctiques i ofereix una alta cobertura.

TERSEO+T2T3 transducer (Saquete, 2010) és un altre enfocament basat en regles de processament de *timex*. Aquest va ser construït per a adaptar unes regles del TIMEX2 al TIMEX3.

USFD2 (Derczynski & Gaizauskas, 2010) abordà el processament de *timex* i de relacions temporals, i es basa en regles morfosintàctiques fetes a mà. També posa en relleu la importància dels assenyaladors temporals en la categorització de les relacions temporals.

Edinburgh-LTG (Grover, Tobin, Alex, & Byrne, 2010) està dirigit al processament de *timex* i esdeveniments. Aquest és també un sistema basat en regles que utilitza el coneixement morfosintàctic, a excepció de processament d'esdeveniments, on la semàntica lèxica s'utilitza per a detectar les nominalitzacions. Edimburgh-LTG utilitza un lèxic d'esdeveniments derivats de les dades d'entrenament i hipònims de WordNet dels conceptes *esdeveniment* i *estat*.

KUL (Kolomiyets & Moens, 2010) és dirigit a *timex* i utilitza un enfocament híbrid. Aquest sistema utilitza un classificador de màxima entropia

¹¹La nostra proposta, TIPSem, s'exclou perquè es detalla en el capítol 3.

per al reconeixement de *timex* i regles manuals per a la tasca de normalització. La novetat de l'enfocament resideix en l'ampliació dels exemples de corpus amb un model de llenguatge de paraula latent (LWLM) i sinònims de WordNet.

JU_CSE_TEMP (Kolya, Ekbal, & Bandyopadhyay, 2010) és un sistema híbrid en fase de desenvolupament que va abordar totes les tasques. Aquest utilitza les regles manuals sobre la morfosintaxi per al processament de *timex* i esdeveniments, i per a abordar les relacions temporals usa CRFs amb un conjunt de característiques basades en els atributs dels *timex* i esdeveniments.

NCSU (Ha, Baikadi, Licata, & Lester, 2010) només va abordar la tasca de categorització de relacions temporals amb una estratègia basada en aprenentatge automàtic (lògica de Markov). A part de les característiques morfosintàctiques, també inclou algunes característiques basades en la semàntica lèxica.

TRIOS (UzZaman & Allen, 2010) és un enfocament híbrid que utilitza la informació morfosintàctica i una representació lògica-semàntica del text basada en rols semàntics. A més de les regles fetes a mà, utilitza CRFs i xarxes lògiques de Markov com a tècniques de ML per a construir models. Va abordar totes les tasques en anglès.

Els detalls sobre el rendiment dels sistemes que van participar en les avaluacions TempEval-1 i TempEval-2 poden consultar-se en l'Apèndix D.

Anàlisi comparativa

Per tal de resumir l'estat de la tècnica, s'analitzen els actuals sistemes de processament de la informació temporal pel que fa a la seua estratègia, les entitats temporal cobertes, i utilitza els coneixements lingüístics.

- **Estratègia.** Les propostes es divideixen en (i) les que utilitzen regles basades en el coneixement de la tasca dels desenvolupadors¹² (estratègia rule-based/racionalista), (ii) les que es basen en la probabilitat o aprenentatge automàtic (estratègia data-driven/empirista), i (iii) les que posen en pràctica una estratègia híbrida.
- **Entitats temporals cobertes.** Les propostes poden cobrir una, diverses o totes aquestes entitats: *timex*, esdeveniments, i relacions temporals.
- **Coneixement lingüístic emprat.** Les propostes poden incloure: (i) coneixement *morfosintàctic* (*ms*), que inclou la segmentació de frases,

¹²patrons, expressions regulars, i regles.

CAPÍTOL 2. ESTAT DE LA QÜESTIÓ

la tokenització, la lematització, l'etiquetatge de categories gramaticals (PoS tagging), l'anàlisi sintàctica, i les llistes de paraules clau¹³; (ii) *semàntica lèxica* (*ls*), que inclou la semàntica en l'àmbit de paraula; i (iii) *rols semàntics* (*sr*), que inclou la semàntica en l'àmbit de frase.

Seguint la tipologia d'anàlisi descrita, la taula 2.4 resumeix els sistemes més recents i destacats.

| Strategy | System | Entities ^a | Linguistic Knowledge ^b |
|--------------------|--|-----------------------|-----------------------------------|
| Rule-based | WVALI (Puscasu, 2007) | trel | ms |
| | XRCE-T (Hagège & Tannier, 2007) | trel | ms |
| | HeidelTime (Strötgen & Gertz, 2010) | timex | ms |
| | USFD2 (Derczynski & Gaizauskas, 2010) | timex,trel | ms |
| | TERSEO (Saquete, 2010) | timex | ms |
| | Edinburgh-LTG (Grover et al., 2010) | timex,event | ms, ls |
| Data-driven | (Boguraev & Ando, 2005) | all | ms |
| | STEP (Bethard & Martin, 2006) | event | ms, ls |
| | (Mani et al., 2006) | trel | ms |
| | (Lapata & Lascarides, 2006) | trel | ms, ls |
| | CU-TMP (Bethard & Martin, 2007) | trel | ms |
| | USDF (Hepple et al., 2007) | trel | ms |
| | NAIST (Cheng et al., 2007) | trel | ms |
| | (March & Baldwin, 2008) | event | ms |
| | NCSU (Ha et al., 2010) | trel | ms, ls |
| Hybrid | TTK ^c (Verhagen et al., 2005) | all | ms, ls |
| | LCC-TE (Min et al., 2007) | trel | ms,ls |
| | KUL (Kolomiyets & Moens, 2010) | timex | ms, ls |
| | JU_CSE_TEMP (Kolya et al., 2010) | all | ms |
| | TRIOS (UzZaman & Allen, 2010) | all | ms,ls,sr |

^atrel (temporal relations)

^bms (morphosyntactic), ls (lexical semantics), sr (semantic roles)

^cTTK is mainly rule-based but uses statistical techniques for event

Taula 2.4: Aproximacions de processament TimeML per a l'anglès

Els sistemes basats en regles són molt populars entre els que aborden el processament TIMEX només (per exemple, HeidelTime i TERSEO+T2T3). Això és probablement perquè les regles són més adequades per a les tasques on es requereix la normalització.

Per contra, els sistemes basats en dades (ML) o híbrids són més populars entre els sistemes que tracten el processament d'esdeveniments i de la relacions temporals. Atès que hi ha una àmplia varietat de mitjans del

¹³Una llista predefinida de paraules que es probable que apareguen en una *timex* o esdeveniment (p. e.: *year*, *war*)

llenguatge que poden expressar aquestes entitats, és més difícil capturar-les usant enfocaments basats en regles.

Pel que fa als coneixements lingüístics utilitzats, la majoria dels sistemes es basen únicament en la morfosintaxi. Alguns d'aquests inclouen la semàntica lèxica i només un aplica els rols semàntics (TRIOS). Això indica que (i) les propietats morfosintàctiques de la llengua són crucials per a fer front al processament de la informació temporal i (ii) l'ús de la semàntica per a aquesta tasca, en particular, els rols semàntics, és molt recent i els seus beneficis sobre la morfosintaxi encara han de ser analitzats (un dels objectius d'aquesta tesi).

Aproximacions per a altres llengües

A més d'anglès, TempEval-2 inclou corpus en cinc idiomes. No obstant això, només hi va haver participants d'anglès i espanyol. A més de la nostra proposta (TIPSem) que es descriu en el capítol 3, només un sistema va participar per a l'espanyol: UC3M (Vicente-Díez, Moreno-Schneider, & Martínez, 2010). Aquest només es dirigia al processament d'expressions temporals. El sistema de la UC3M, es basa en un conjunt de regles morfosintàctiques obtingudes centrant-se en les expressions temporals més freqüents en espanyol. Fora del TempEval, hi ha contribucions per a altres idiomes.

Per al xinès, una proposta primerenca en relació amb el processament de la informació temporal va ser presentada per Li et al. (2001). Posteriorment, els mateixos autors es van dirigir a la tasca de categorització de relacions temporals amb dues tècniques de ML - arbres de decisió i classificadors bayesians (Li, Wong, Cao, & Yuan, 2004). Hi ha alguns mètodes per al reconeixement de *timex*, però cap amb TimeML. Hacioglu et al. (Hacioglu et al., 2005) va avaluar un enfocament basat en ML usant màquines de suport vectorial. Aquestes propostes es refereixen a un esquema previ a TimeML i per tant els resultats no són comparables. Zhang et al. (Zhang et al., 2008) utilitza una tècnica basada en l'error. Aquest enfocament basat en regles presentà el seu propi esquema d'anotació temporal (no TimeML) i el seu propi corpus anotat a causa de la manca de recursos, cosa que dificulta les possibilitats de comparació. Es va assenyalar que les causes principals d'error per al reconeixement de *timex* són les ambigüitats morfosintàctiques. Val la pena esmentar que no existeixen mètodes per al processament d'esdeveniments en xinès.

Per a altres idiomes, ens trobem amb algunes propostes: sistemes associats amb les extensions del TimeML a idiomes, com es descriu en la secció 2.2. Per exemple, per a l'italià (Caselli, dell'Orletta, & Prodanof, 2009; Robaldo,

Caselli, Russo, & Grella, 2011) i francès (Bittar, 2009). També val la pena esmentar els esforços multilingües de l'esquema d'anotació TIDES (limitat a expressions temporals) per als idiomes àrab i xinès al TERN (2004), i per a l'italià en EVALITA'07¹⁴.

En resum, es pot observar que la majoria dels sistemes es basen en la morfosintaxi i l'ús de la semàntica és molt limitada. Només uns pocs sistemes utilitzen la semàntica lèxica i només un d'aquests utilitza els rols semàntics.

2.4 L'ús de la semàntica

A la secció anterior, hem vist que només uns pocs sistemes de l'estat de la qüestió s'acosten a l'ús d'informació semàntica per al processament de la informació temporal. Aquesta tesi explora l'aplicació de dos tipus d'informació semàntica: (i) **semàntica lèxica** basada en WordNet (Fellbaum, 1998), i (ii) **rols semàntics** sobre el conjunt de rols PropBank (Palmer et al., 2005). En aquesta secció, presentem els fonaments teòrics i computacionals.

2.4.1 Semàntica lèxica

La semàntica lèxica estudia el significat de les paraules –més precisament, dels lexemes– i les relacions semàntiques entre aquestes, com ara sinonímia, antonímia, hiperonímia, i hiponímia (Lyons, 1981). Des d'un punt de vista computacional, l'objectiu de la semàntica lèxica és el modelatge automàtic del significat dels lexemes: representa el significat de paraules, sentits de les paraules, i les relacions semàntiques entre aquestes. Les relacions semàntiques principals són les relacions de sinonímia (paraules amb el mateix significat), i les relacions jeràrquiques (hiperonímia i hiponímia). Les relacions jeràrquiques representen un sentit específic que està inclòs en un sentit més general (hiperonímia) i viceversa (hiponímia).

Com ja hem vist, molts dels enfocaments proposats es basen en els coneixements morfosintàctics, especialment en les tasques de processament de *timex*. En particular, la inclusió de llistes de paraules clau probables d'aparèixer en *timex* o esdeveniments ha demostrat ser útil per a aquesta tasca. Des del nostre punt de vista, l'aplicació de llistes de paraules clau podria ser considerada com una forma de semàntica lèxica de domini tancat, ja que són específiques per a les *timex* o esdeveniments, i aquests no reflecteixen els diferents sentits, ni les relacions semàntiques (per exemple, hiperonímia).

¹⁴<http://www.evalita.it/2007/tasks/tern>

L'aplicació dels recursos generals de semàntica lèxica com xarxes semàntiques introdueix dos avantatges respecte a les llistes de paraules clau. En primer lloc, un recurs general de semàntica lèxica captura no només la semàntica lèxica d'una paraula (com una seqüència de caràcters) en un domini específic (per exemple, el temps/eventualitat), sinó la semàntica de la paraula (com un concepte específic o sentit) en un domini general, codificat en un lèxic amb una estructura de xarxa semàntica. En segon lloc, les xarxes semàntiques generals són estàndard, completes, i són mantingudes i millorades per tercers, independentment dels seus usos particulars. Per tant, l'ús de la semàntica lèxica augmenta la modularitat dels enfocaments i evita la costosa tasca de desenvolupar una llista completa de paraules clau.

WordNet i EuroWordNet (Vossen, 1998) representen les xarxes semàntiques més populars per a l'anglès i idiomes europeus, respectivament. Es tracta de grans bases de dades lèxiques. Substantius, verbs, adjectius i adverbis s'agrupen en conjunts de sinònims (synsets), cadascun expressant un concepte diferent. Els synsets estan vinculats entre si en una xarxa per mitjà de les relacions conceptuals-semàntiques i lèxiques (p. e.: sinonímia, hiperonímia).

En el processament de *timex*, les xarxes semàntiques han estat aplicades per dues propostes. Negri et al. (Negri & Marseglia, 2004) utilitzà WordNet per crear una llista d'entitats amb nom temporal com *Bastille Day*, *Hanukkah*, etc., mitjançant la recopilació de tots els hipònims del synset *calendar_day*. Saquete et al. (2004) utilitzà xarxes semàntiques per expandir una llista de paraules clau amb els seus sinònims. Això demostra que la informació continguda en les xarxes semàntiques pot ser útil per a l'extracció d'expressions temporals.

En el processament d'esdeveniments, trobem tres propostes d'ús de WordNet (Saurí et al., 2005; Bethard & Martin, 2006; Grover et al., 2010). Aquestes deriven de les llistes de paraules clau dels hipònims de *esdeveniment* i *estat*. Per tant, la informació de les xarxes semàntiques ha demostrat ser útil també per a aquesta tasca.

Finalment, en relació a la categorització de relacions temporals la semàntica de les xarxes semàntiques no ha estat utilitzada directament (Verhagen et al., 2007). WordNet només s'usava per a ampliar la informació dels participants en la relació, en particular l'extensió de les classes dels esdeveniments TimeML amb les classes de l'ontologia superior de WordNet.

Segons Lyons (1981), la semàntica lèxica no es pot considerar de forma aïllada. El significat dels lexemes depèn del significat de les oracions en què tenen lloc, i el significat d'una oració depèn del significat dels seus lexemes constituents. Per tant, la semàntica en l'àmbit de l'oració (rols semàntics, per exemple) es crucial per a captar el significat.

2.4.2 Rols semàntics

Els rols semàntics capten el significat d'una frase pel que fa a com es relacionen els arguments dels seus predicats. Un rol semàntic és la relació conceptual entre un component i el seu predicat. Per a un predicat, cada component és un argument (agent, pacient, etc.) o un complement (locatiu, temporal, etc.) L'exemple (28) mostra frases marcades amb rols semàntics indicats entre claudàtors.

- (28) [In July 1277 *Temporal*], [John *Agent*] wrote [a letter *Patient*] [with a quill *Manner*]
[in his chamber *Locative*].

En (28), en relació amb el verb que governa el predicat (*to write*), *TEMPORAL* significa l'adjunt temporal (*quan* va ocórrer), *AGENT* representa l'argument agent (escriptor/autor), *PATIENT* és l'argument pacient (cosa escrita/text), *MANNER* és l'adjunt de manera (*com* es va fer), i *LOCATIVE* és l'adjunt locatiu (*on*).

Els orígens dels rols semàntics són els *karakas* del gramàtic del sànscrit Panini (Kiparsky, 2002). Des d'aleshores, s'han dedicat molts anys a l'estudi dels rols. Les primeres propostes modernes es remunten als anys seixanta (Gruber, 1965; Fillmore, 1968). A partir d'aquestes propostes van sorgir un gran nombre de jocs de rols semàntics. Les propostes van des de molt específics amb molts rols (p. e.: *menjador*, *cosa menjada*) a molt abstractes amb pocs rols (p. e.: *agent*, *temporal*). Els diferents conjunts de rols poden ser definits en una jerarquia dels més detallats (p. e.: *menjador*, etc.) a només dos *proto-rols*: proto-agent i proto-pacient (Dowty, 1991).

Avui dia, en el PLN, els jocs més utilitzats són el que es van desenvolupar en el projecte FrameNet (FN) (Johnson et al., 2002; Ruppenhofer et al., 2005) i el definit en el projecte PropBank (PB) (Palmer et al., 2005). Cada proposta se centra en un nivell diferent de granularitat.

D'una banda, FN defineix una representació detallada de situacions, incloent-hi un conjunt de rols molt específiques – vegeu ((29).

- (29) [In July 1277 *Time*], [John *Author*] wrote [a letter *Text*] [with a quill *Manner*]
[in his chamber *Place*].

D'altra banda, PB consisteix en un conjunt limitat de rols abstractes – vegeu (30).

- (30) [In July 1277 $AM-TMP$], [John A_0] wrote [a letter A_1] [with a quill $AM-MNR$]
[in his chamber $AM-LOC$].

FN evita el problema de la definició d'un conjunt petit de rols abstractes amb la definició de tants rols com siguin necessaris amb una pèrdua d'informació mínima. No obstant això, el conjunt de PB ofereix més cobertura lèxica que FN, que només cobreix els verbs més generals de l'anglès i alguns substantius.

La diferència en la cobertura lèxica de FN i el PP deriva principalment de seus enfocaments: PB ha anotat tots els verbs en una secció de *The Wall Street Journal*, mentre que FN ha treballat en l'anglès més general. A més, és més fàcil obtenir més cobertura amb una representació abstracta d'un conjunt de rols petit i limitat (PB) que amb un conjunt de rols específics gran (FN), que ha de ser estès per a cada nou verb que aparega en el corpus.

Juntament amb l'estudi teòric dels rols semàntics, els investigadors estudien els mètodes computacionals per al seu etiquetatge automàtic en el text. L'etiquetatge de rols semàntics (SRL) ha aconseguit importants resultats (Gildea & Jurafsky, 2002; Punyakanok, Roth, Yih, Zimak, & Tu, 2004; Moreda, 2008). Molts estudis sobre l'aplicació dels rols semàntics han demostrat que aquesta informació és útil per a diferents propòsits, com ara el resum automàtic de textos (Melli et al., 2006), la recuperació d'informació (Moreda et al., 2007), i la cerca de respostes (Moreda et al., 2011).

Centrant-nos en l'aplicació dels rols semàntics en el processament de la informació temporal, només es troba una proposta: TRIOS (UzZaman & Allen, 2010). TRIOS utilitza una representació lògica dels rols semàntics sobre la base d'una variació de FrameNet. La variació redueix el conjunt de rols FrameNet a aproximadament 30 rols semàntics.

En treballs relacionats, la influència dels rols semàntics en la tasca no ha estat analitzada. A més, encara que proporcione informació útil per a interpretar les entitats temporals, el conjunt de rols de PropBank mai s'ha utilitzat per aquest propòsit. L'exemple (31) mostra frases marcades amb rols de PropBank entre claudàtors i les entitats temporals subratllades.

- (31) a. [In July 1277 $AM-TMP$], [John A_0] wrote [a letter A_1] [with a quill $AM-MNR$]
[in his chamber $AM-LOC$].
b. [The tax rates A_1] were [lower A_2].
c. [The company A_0] completed [the transaction A_1] [yesterday $AM-TMP$].
d. [Iraq A_0] invaded [Kuwait A_1] [because of the disputes over oil $AM-CAU$].
e. [He A_0] left [the party A_1] [after [she A_0] kissed [John A_1] $AM-TMP$].

Com es pot veure, les expressions temporals tendeixen a ser contingudes per adjunts temporals. En casos especials, les *timex* també poden trobar-se en arguments numerats de verbs que requereixen un argument temporal. Per exemple, *to last* requereix un període de temps, que es representa en PropBank amb l'argument numerat A2.

Els esdeveniments són normalment representats pels verbs principals. A més, l'argument A2 del verb *to be* conté sovint esdeveniments estat (31b). Els esdeveniments no verbals tenen més probabilitats d'aparèixer en complements específics (31d), o en el rol A1 de certs verbs (31c).

La subordinació de les relacions temporals, sovint es denota per subordinacions temporals marcades per complements temporals com en (31e). En aquests casos, la conjunció que governa la subordinació, en l'exemple *after*, aporta informació rellevant sobre el tipus de la relació.

2.5 Conclusions

Aquesta tesi se centra en la tasca de processament de la informació temporal, que s'emmarca en l'àrea de processament del llenguatge natural (NLP). El processament de la informació temporal consisteix en la interpretació automàtica i el raonament sobre els aspectes temporals dels textos, que és crucial per a altres aplicacions del PLN com la recuperació de la informació i la cerca de respostes.

La investigació sobre el processament de la informació temporal ha evolucionat a partir de diferents estratègies racionalistes a l'estratègia empírica basada en corpus. Aquesta última consisteix en l'anotació de corpus seguint un esquema d'anotació temporal. Actualment, l'esquema estàndard d'anotació temporal és TimeML, a causa de la seua integritat i les millores que presenta sobre els seus predecessors. TimeML capta les expressions lingüístiques que transmeten esdeveniments, el temps i les relacions temporals.

L'anotació automàtica de TimeML és una tasca complexa, que requereix coneixements lingüístics de caràcter semàntic, del discurs i pragmàtics. Els esdeveniments, el temps i les relacions temporals poden ser ambigües des del punt de vista morfosintàctic. No obstant això, la majoria dels plantejaments computacionals de l'estat de la qüestió per a l'anotació TimeML automàtica només utilitzen la informació morfosintàctica. Les escasses propostes que inclouen la semàntica se centren en l'ús de la semàntica lèxica i només una inclou els rols semàntics (rols FrameNet). D'altra banda, la influència de la semàntica no ha estat analitzada en profunditat i, en particular, el conjunt de rols PropBank mai s'ha utilitzat per al processament de la informació temporal.

El capítol següent descriu el nostre enfocament semàntic per al processament de la informació temporal: TIPSem. La proposta difereix dels enfocaments l'estat de la tècnica en els aspectes següents: TIPSem aplica la semàntica lèxica i rols semàntics, i analitza en particular la seua contribució a aquesta tasca; el nostre enfocament utilitza el conjunt de rols PropBank, que proporciona una informació valuosa sobre les entitats temporals, però que mai s'ha utilitzat per a aquest propòsit; i mentre que tots els enfocaments de l'estat de la tècnica es desenvolupen i s'analitzen per a un idioma, TIPSem s'aplica i s'analitza de forma multilingüe en anglès, espanyol, italià i xinès sobre el corpus públic del TempEval-2.

Capítol 3

TIPSem: una aproximació semàntica al processament de la informació temporal

En aquesta secció es descriu la nostra proposta per al tractament automatitzat de la informació temporal. Com que el seu element central és la semàntica, hem anomenat la nostra proposta **TIPSem**, acrònim de l'anglès *Temporal Information Processing using Semantics*. TIPSem és un enfocament híbrid (basat en aprenentatge automàtic i basat en regles), que explota la semàntica lèxica (xarxes semàntiques) i els rols semàntics, a més de la morfosintaxi. La informació temporal es processa i es representa seguint l'esquema d'anotació TimeML.

L'arquitectura de TIPSem es divideix en tres mòduls principals que corresponen a les subtasques de processament dels principals elements TimeML:

- **Processament d'expressions temporals (etiqueta TimeML TIMEX).** Aquesta tasca implica el reconeixement¹ en el text, la classificació, i la normalització a l'estàndard ISO 8601 d'aquestes entitats.
- **Processament d'esdeveniments (etiqueta TimeML EVENT).** Aquesta tasca implica el reconeixement en el text, i la classificació d'aquestes entitats.
- **Categorització de relacions temporals (etiqueta TimeML TLINK).** Aquesta tasca implica la categorització dels diferents tipus de relacions entre les entitats anteriors. És a dir, les relacions entre els es-

¹Delimitació exacta en el text.

deveniments i *timex*, entre els esdeveniments i la data de creació del document, entre els esdeveniments subordinats a la mateixa frase (intraoracional), i entre els principals esdeveniments de diferents frases (interoracional).

Aquest capítol està estructurat de la manera següent. La primera secció mostra l'arquitectura general de l'enfocament. La segona secció descriu en detall les parts de l'enfocament i mostra com el processament de cada entitat temporal es realitza amb la semàntica. La tercera secció analitza les tècniques d'aprenentatge màquina emprades. Finalment, la darrera secció detalla l'extensió de l'aproximació a altres llengües: espanyol, italià i xinès.

3.1 Arquitectura general

TIPSem és un enfocament híbrid. Està principalment basat en aprenentatge automàtic, però també inclou una base de regles basades en el coneixement temporal: TIMEK². La figura 3.1 il·lustra l'arquitectura del sistema TIPSem i en mostra les seues diferents parts.

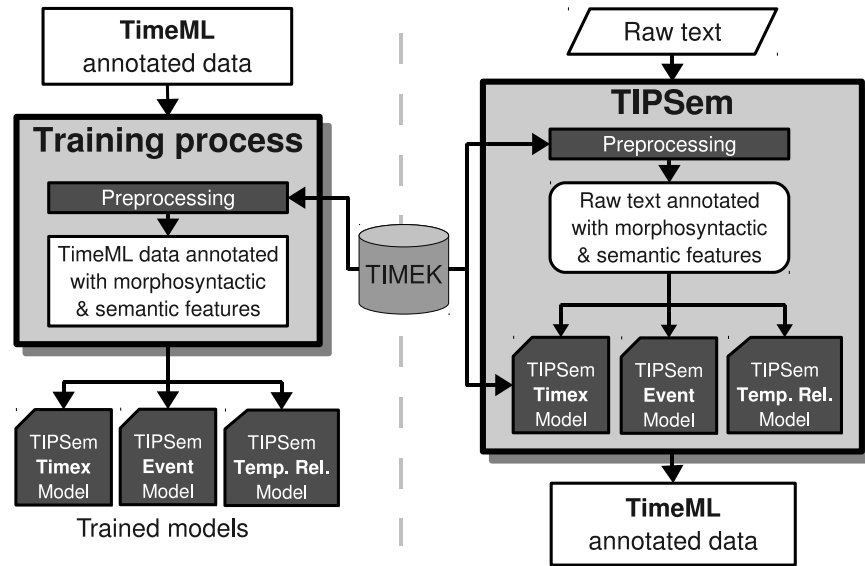


Figura 3.1: Arquitectura del TIPSem

²TIMEK es detalla en l'apèndix A.

Com un enfocament principalment basat en aprenentatge automàtic, consta de dos processos: el procés d'entrenament i el procés d'aplicació.

En tots dos processos, el primer pas és el preprocessament de l'entrada. Aquest consisteix en l'obtenció de les característiques següents:

- **Característiques morfosintàctiques.** Aquestes són les paraules, lemes, i categories gramaticals (PoS), les quals s'obtenen mitjançant TreeTagger (Schmid, 1994), i l'arbre sintàctic (estil Treebank) obtingut usant l'analitzador de Charniak (Charniak & Johnson, 2005).
- **Característiques semàntiques.** La principal característica del nostre enfocament és l'ús de trets semàntics per a millorar el rendiment dels enfocaments basats només en la morfosintaxi. Aquests es divideixen en tres grups:
 - *Semàntica lèxica (LS).* Aquestes inclouen la informació d'hiperonímia obtinguda de WordNet (Fellbaum, 1998) i TIMEK. La semàntica lèxica captura la relació de les paraules amb significats temporals o eventius en l'àmbit lèxic.
 - *Rols semàntics (SR).* Aquestes inclouen l'etiquetatge de rols semàntics de PropBank obtingut usant l'eina CCG (Punyakanok et al., 2004). Els rols semàntics capturen la relació de les paraules amb el rol que exerceixen en la frase (o més precisament, en el predicat).
 - *Semàntica temporal (TS)*³. Aquestes inclouen les propietats semàntiques de les *timex* (comparació dels valors) obtingudes del TIMEK.

En primer lloc, el procés d'entrenament obté models per a cadascuna de les tasques prèviament definides a partir de les dades TimeML preprocessat. Aquest procés es realitza una sola vegada.

En segon lloc, el procés d'aplicació aplica els models per a obtenir l'anotació TimeML de qualsevol entrada de text preprocessat.

³Només utilitzada per a la categorització de relacions temporals.

Com a part de l'arquitectura basada en aprenentatge automàtic, el mòdul basat en regles consisteix en una base de coneixements temporals inclosos en el sistema: TIMEK. Aquesta s'utilitza en el preprocessament per a obtenir algunes característiques LS sobre les dades d'entrada, que s'utilitzen en la classificació de *timex* i la normalització. També en el preprocessament, s'utilitza per a obtenir les funcions de TS que utilitza en relació amb la categorització temporal. Finalment, TIMEK s'utilitza en el processament de *timex* per a normalitzar la *timex* a l'estàndard ISO 8601.

Un cas típic de l'us de TIPSem s'il·lustra en (32). Aquest mostra un text d'entrada en anglès i l'eixida TimeML obtinguda per TIPSem.

(32) Document creation time (DCT): 2010-06-29

Input: John came on Monday. After, he left the US. Then, he told us he was born in 1980.

Output:

```
<?xml version="1.0"?> <TimeML>
<TIMEX3 tid=t0 type="DATE" value="2010-06-29" temporalFunction="false"
functionInDocument=CREATION_TIME>>2010-06-29</TIMEX3>

<TEXT>
John <event class="OCCURRENCE" eid="e1">came</event>
on <timex3 type="DATE" value="2010-06-28" tid="t1">Monday</timex3>.
After, he <event class="OCCURRENCE" eid="e2">left</event> the US.
Then, he <event class="REPORTING" eid="e3">told</event> us
he was <event class="OCCURRENCE" eid="e4">born</event> in
<timex3 type="DATE" value="1980" tid="t2">1980</timex3>.
</TEXT>

<TLINK lid=1 reltype="during" eventID="e1" relatedToTime="t1"/>
<TLINK lid=2 reltype="after" eventID="e2" relatedToEvent="e1"/>
<TLINK lid=3 reltype="after" eventID="e3" relatedToEvent="e2"/>
<TLINK lid=4 reltype="during" eventID="e4" relatedToTime="t2"/>
</TimeML>
```

Altres exemples poden ser provats utilitzant la demostració en línia del sistema de TIPSem disponible en <http://gplsi.dlsi.ua.es/demos/TIMEE/>.

3.2 Arquitectura específica de cada tasca

Les subseccions següents descriuen les parts principals del sistema respecte a cada mòdul de l'enfocament TIPSem.

3.2.1 Processament d'expressions temporals (*timex*)

Com es descriu en la secció 2.2, en TimeML, una *timex* és una representació lingüística d'un punt en el temps o període. Les *timex* es poden classificar en els següents tipus: data, hora, durada i repetició; i s'han de normalitzar al seu valor ISO 8601. Per tant, el processament computacional de *timex* en llenguatge natural es compon de tres subtasques principals: (i) el reconeixement d'aquestes expressions en el text, (ii) la seua classificació, i (iii) l'obtenció del seu valor normalitzat.

L'exemple (33) il·lustra un text pla d'entrada i l'eixida del sistema esperada.

(33) Document creation time (DCT): 2010-06-09)

Input: John came on Tuesday. He will stay here for one year and two months.

Output:

John came on <timex3 type="DATE" value="2010-06-08">Tuesday</timex3>.
He will stay here for <timex3 type="DURATION" value="P1Y2M">one year
and two months</timex3>.

La figura 3.2 resumeix l'arquitectura de la part del processament de *timex*. Aquest mostra la proposta (TIPSem) i els recursos que aquesta utilitza.

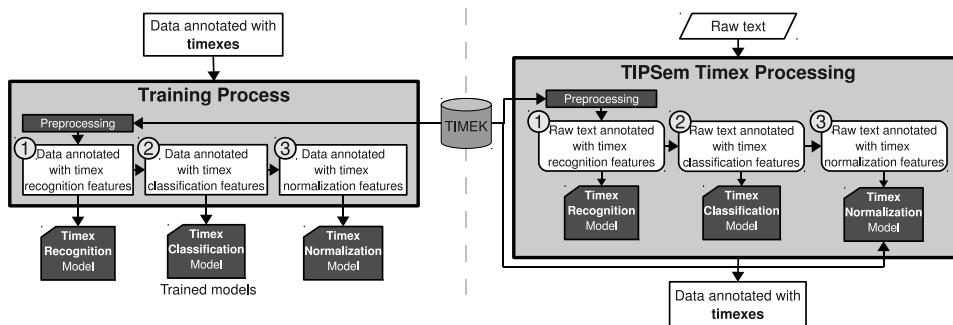


Figura 3.2: TIPSem arquitectura de processament de *timex*

De manera anàloga a l'arquitectura general, el primer pas de tots dos processos d'entrenament i d'aplicació és el preprocessament de l'entrada.

A l'entrenament, els models es generen de manera seqüencial en tres passos. En primer lloc, les característiques de reconeixement s'obtenen i el model es genera. En segon lloc, les característiques per a la classificació s'obtenen de les funcions de reconeixement i el model de classificació s'obté. Finalment, les característiques de la normalització s'obtenen a partir dels criteris de classificació i el model de normalització s'obté.

En el procés d'aplicació, TIPSem aplica el model de reconeixement, després aplica el model de classificació per a la *timex* reconeguda, i finalment aplica el model de la normalització a les *timex* classificats. Després d'aplicar el model de normalització, TIMEK s'utilitza per a obtenir la normalització ISO 8601.

Reconeixement de *timex*

L'aprenentatge automàtic és útil en problemes en què la complexitat impedeix el modelatge manual de la tasca. El problema d'anotació automàtica TimeML pot abordar-se amb tècniques d'aprenentatge supervisat, sempre que hi ha corporacions que contenen un nombre representatiu d'exemples anotats (per exemple, les dades del TempEval-2). Les primeres versions de TIPSem estaven completament basades en regles (Llorens et al., 2009, 2009c)⁴. No obstant això, la construcció manual de regles de cobertura alta era costós.

La versió de TIPSem presentada en aquesta tesi per al reconeixement de *timex* és completament dirigida per les dades. Ens ocupem del problema que planteja aquesta tasca com un problema d'etiquetatge de seqüència, que també pot ser vist com un problema de classificació. En aquesta tasca de delimitació, s'utilitzen etiquetes IOB2⁵ per a classificar totes les paraules. Tenint en compte un text d'entrada, cada paraula ha de ser classificada com el començament d'una *timex* (B-*timex*), dins d'una *timex* (I-*timex*), o fora d'una *timex* (O). L'alfabet IOB2 resultant consta de B-*timex*, I-*timex* i O. L'exemple (34) lustra el problema de reconeixement de *timex*.

| | | |
|------|-------------------|-----------------|
| (34) | <u>input text</u> | <u>solution</u> |
| | There | O |
| | was | O |
| | no | O |
| | profit | O |
| | this | B- <i>timex</i> |
| | year | I- <i>timex</i> |
| | . | O |

La nostra proposta presenta un conjunt de característiques obtingudes en el nivell de paraula, és a dir, cada paraula té el seu propi conjunt de característiques.

⁴Aquestes regles també es van analitzar per a l'espanyol (Llorens et al., 2009b) i català (Llorens et al., 2009a).

⁵Format IOB2: (B) *begin*, (I) *inside*, and (O) *outside*.

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

Per a atacar el problema plantejat, definim les següents **característiques morfosintàctiques**.

- **Morfològiques:** Aquestes característiques són paraula, lema, i PoS en un context dinàmic de com a molt un finestra de 7 (-3,+3).

Per a algunes *timex* com (35), la paraula i el lema són crucials per al reconeixement.

(35) The conference ended [*yesterday* *TIMEX3*].

Com es pot veure en l'exemple, alguns lemes, com *yesterday*, és molt probable que representen *timex*⁶. No obstant això, l'exemple (36) mostra una expressió temporal que pot ser ambigua en el nivell lèxic, *May*⁷.

(36) *May* you find what you are looking for.
She will be back in [*May* *TIMEX3*].

Ací, la informació del PoS resol l'ambigüitat perquè el primer *May* és un verb modal, mentre que el segon és un nom propi (un mes).

Per a expressions multiparaula, el context té un paper important perquè algunes paraules només formen part d'expressions temporals si les paraules veïnes tenen uns lemes i PoS específics – vegeu l'exemple (37).

(37) This is the chief *of* the company.
Today is the [3rd *of* June *TIMEX3*].

En l'exemple, la paraula *of* només és part d'una expressió temporal si, per exemple, la seua paraula precedent és un ordinal i la paraula següent és un nom propi amb un lema específic (un mes), però no si la paraula precedent és per exemple *chief* i la següent és *the*.

A més, el context actua com a informació de desambiguació contextual. En l'exemple (38), podem veure que un nombre és més probable que

⁶Excepte quan no tenen un significat temporal, com en el títol de la cançó dels Beatles.

⁷La qual també pot ser ambigua en el nivell morfosintàctic si es refereix al nom de persona.

forme part d'una expressió temporal si està precedit per una preposició i no seguit d'un substantiu.

- (38) He owns 1999 properties.
He was born in [1999 *TIMEX*3].

La finestra dinàmica de 7 va obtenir un millor rendiment que altres tipus i mides de finestres contextuals (Llorens et al., 2011a). Es tracta d'un context motivat per la sintaxi que només inclou el context del constituent sintàctic de la paraula que es vol classificar. En (37), el context estàtic en una finestra de 7 de la paraula *of* és [is, the, 3rd, of, June, ., -], mentre que el seu context dinàmic es [-,the, 3rd, of, June, -, -]. El context dinàmic evita que l'algorisme d'aprenentatge automàtic considere, per exemple, que la paraula *is* poguera ser rellevant per a decidir si *of* és part d'una expressió temporal o no.

Tot i que la informació morfològica és molt útil, no sempre és suficient per a distingir les expressions temporals de la resta del text.

- **Sintàctiques:** Les *timex* es troben en determinats tipus de sintagmes: sintagmes nominals (NP), sintagmes preposicionals (PP), sintagmes adjetivals (ADJP), i els sintagmes adverbials (ADVP). Per al reconeixement de *timex*, utilitzem una característica el valor de la qual indica el tipus de sintagma al qual pertany una paraula. Això és útil per a la identificació de les paraules que poden ser part d'una *timex*.

En l'exemple (39a), aquesta característica assenyalava que *stayed* i *during* no poden participar en una *timex* perquè no pertanyen a cap d'aquests tipus de sintagmes.

- (39) a. (S (NP She) (VP stayed
(PP during (NP the second half of the year))). (*timex*)
b. (S (NP She) (VP stayed
(PP behind (NP the wall)))). (*not timex*)

A més, si un NP és governat per un PP, la preposició governant pot ser útil per a incrementar la probabilitat que el NP continga una *timex*. Per tant, açò també s'inclou com a característica. En l'exemple (39), el PP governat per *during* introdueix una *timex* en l'NP que governa (39a), mentre que el PP governat per *behind* no ho fa (39b).

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

No obstant això, els exemples (40a) i (40b) mostren la mateixa estructura sintàctica i només (40a) representa una *timex*.

- (40) a. (S (NP She) (VP won
(PP in (NP April))). (*timex*)
b. (S (NP She) (VP won
(PP in (NP Africa)))). (*not timex*)

Això motiva l'ús d'informació d'un nivell lingüístic més alt, com la semàntica. Per exemple, la semàntica lèxica assenyala que el concepte *April* està relacionat amb el temps i els rols semàntics mostren que *April* exerceix una funció temporal en la frase.

Les **característiques semàntiques** emprades per a millorar el model són:

- **Lèxica semàntica:** Les classes ontològiques més altes de WordNet representen el significat de les paraules de manera ontològica. La majoria dels noms continguts en *timex* són hipònims dels conceptes en anglès ***time***, ***time-period*** o ***time-unit***. Aquests conceptes estan al quart nivell des del concepte més general, *entity*. Per aquesta raó, la nostra aproximació considera com a característica els quatre hiperònims més generals (*top4hypers*) del sentit més comú (*first sense*) de cada nom i verb, excloent-ne el concepte més general (vegeu l'exemple 41).

- (41) *top4hypers* per a *decades* és:

decade: **time-period** => quantity => measure => abstraction

Atès que moltes de les *timex* contenen paraules amb valors relacionats amb el temps per a *top4hypers*, aquesta característica incrementa la probabilitat representar *timex* de les paraules que obtinguen aquests valors, fins i tot si els seus lemes no apareixen en l'entrenament, la qual cosa afavoreix la generalització⁸.

- **Rols semàntics:** Els rols semàntics proporcionen les relacions semàntiques estructurals del predicats en les quals poden participar els elements TimeML. El rol semàntic temporal (principalment ***AM-TMP***⁹) representa arguments i adjunts temporals que moltes vegades contenen

⁸L'ambigüitat lèxica ha sigut solucionada amb el *baseline* (sentit més comú). En treballs futurs, s'hi empraran tècniques de desambiguació del sentit.

⁹En casos específics, alguns arguments numerats de verbs denoten temporalitat (p. e.: A2 de *to last*).

timex. Per a cada paraula d'un predicat, considerem dues característiques: *role* (*aïllat*) i *role + lemma*.

La característica *role* hauria d'incrementar la probabilitat de representar una *timex* de totes les paraules que pertanyen a rols temporals independentment dels seus lemes (i també reduir la probabilitat de les paraules que estiguen fora d'aquests rols).

La combinació *role + lemma* hauria d'incrementar específicament la probabilitat de representar una *timex* de les combinacions de lemes relacionats amb les *timex* (p. e.: *April*, *week*, etc.) i els rols temporals; així com també reduir la probabilitat dels lemes no temporals continguts en rols temporals (p. e.: *meeting*, *war*). L'exemple (42) mostra com es pot solucionar l'ambigüitat morfosintàctica amb aquestes característiques.

- | | | |
|------|--|--------------------|
| (42) | [April _{A0}] likes apples. | (nom propi femení) |
| | I went to Canada [in April _{AM-TMP}]. | (mes, TIMEX3) |
| | I went to Canada [after the meeting _{AM-TMP}]. | (esdeveniment) |
| | The Iraqis have resisted attempts to inspect [such quarters _{AM-LOC}]. | (lloc) |
| | [The noun Monday _{A0}] is derived from Middle English <i>Monenday</i> . | (no temporal) |
| | [Last week _{A0}] is my favorite song. | (títol de cançó) |

Les dues primeres frases contenen el nom *April*. No obstant això, en la primera *April* és un nom de persona i en la segona és un mes i per tant una TIMEX3. Quant a rols semàntics, l'expressió *April* representa un rol numerat del verb *to like* en la primera frase, mentre que representa un rol temporal en la segona. La tercera frase, mostra per què la combinació *role + lemma* (*meeting+AM-TMP*) és útil per a solucionar els casos on el rol temporal no representa una *timex*. A més, en la quarta frase, la presència d'un rol locatiu evita considerar *quarters* com a expressió temporal. Finalment, les expressions *Monday* i *Last week* poden no representar *timex*, com es mostra a l'exemple. Els rols no temporals assenyalen quines expressions és menys probable que representen *timex* (p. e.: A0). En resum, la nostra aproximació aprèn dels exemples anotats quins rols semàntics és més possible que continguin expressions temporals.

La figura 3.3 il·lustra el vector de característiques del reconeixement de *timex*.

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

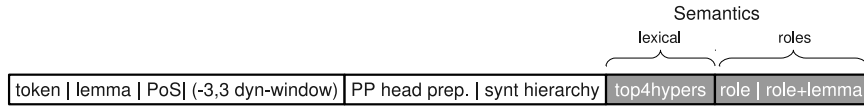


Figura 3.3: TIPSem característiques per al reconeixement de *timex*

Considerant la frase (43), l'exemple (44) mostra els valors d'aquest vector obtinguts per a la paraula *days*.

(43) Bush can call up to 200,000 reservists for up to 180 days without seeking congressional approval.

(44)

| | | | | |
|------|--|------|--|---|
| up | | up | | IN |
| to | | to | | TO |
| 180 | | card | | CD |
| days | | day | | NNS for PP time_unit > measure > abstraction AM-TMP AM-TMP+day |
| - | | - | | - |
| - | | - | | - |
| - | | - | | - |

Classificació de *timex*

Com s'ha descrit en la secció 2.2, hi ha quatre tipus o classes d'expressions temporals: data, hora, duració i repetició. El problema de la classificació consisteix a seleccionar un dels tipus per a cada *timex* reconeguda (vegeu l'exemple 45).

Els tipus *date* i *time* representen referències temporals de diferents granularitats. Les dates inclouen les referències expressades amb unitats temporals majors o iguals que els dies (p. e.: *days*, *months*, *years*, *centuries*, etc.), mentre que les hores inclouen referències expressades amb unitats temporals menors que els dies (p. e.: *hours*, *minutes*, etc.). Per exemple, *yesterday* i *October 1999* són dates, mentre que *yesterday morning* i *8 p.m.* són hores.

El tipus *duration* representa períodes o expressions duratives (p. e.: *two years*, *10-hour*, etc.); i *set* representa expressions recurrents en el temps que defineixen patrons de repetició en el temps (p. e.: *weekly*, *each month*, etc.).

| | | |
|------|--------------|-----------------|
| (45) | <u>TIMEX</u> | <u>solution</u> |
| | 1999 | date |
| | 8 a.m. | time |
| | monthly | set |
| | two years | duration |
| | next Monday | date |

L'aproximació de classificació utilitza les mateixes característiques emprades en el reconeixement, excepte els rols. No obstant això, atès que en la classificació tenim un conjunt de *timex* en compte d'un conjunt de paraules, les característiques definides per al reconeixement han de ser adaptades. Per a convertir característiques basades en paraules en característiques basades en elements, agrupem els valors de la manera següent. Per a *timex* d'una sola paraula (p. e.: *yesterday*), les característiques són equivalents. Per a *timex* de més d'una paraula com *two years* en l'exemple (46a), els valors de les característiques de cada paraula es concatenen amb `_` (p. e.: lemma=two_year,PoS=CD_NNS), i si els valors són compartits per totes les paraules (p. e.: heading-preposition(two)=for, heading-preposition(years)=for) s'agrupen en un sol valor (p. e.: heading-preposition(two_years)=for).

Les seqüències de lemes i PoS sovint representen classes diferents de *timex* i, per tant, són útils per a diferenciar-les. No obstant això, de vegades la mateixa seqüència representa diferents classes, com es mostra en (46). En aquest cas, les característiques sintàctiques, com per exemple les preposicions que governen sintagmes preposicionals (p. e.: *for*, *in*), són útils.

- (46) a. He will be here for [two years *DURATION*].
b. He will be here within [two years *DATE*].

En la classificació de *timex*, els rols semàntics són massa generals per a distingir-ne les classes, perquè totes les *timex* estan contingudes en el mateix tipus de rols, principalment temporals.

A més de les característiques incloses en el reconeixement, hi hem afegit característiques específiques a la classificació de *timex*. Aquestes són obtingudes principalment usant TIMEK, el qual reforça la semàntica lèxica de l'aproximació. Les característiques afegides són les següents:

- **Semàntica lèxica (TIMEK¹⁰):**

- **Patró:** Per a cada *timex* reconeguda, el seu patró es construeix seguint les especificacions del patró TIMEX definides en l'apèndix A. Amb el patró, l'algorisme d'aprenentatge automàtic pot aprendre models més generals a partir dels patrons en compte dels lemes o PoS. Per exemple, les *timex two weeks ago* i *100 days ago* obtindrien el mateix patró: *NUM TUNIT ago*.
- **Granularitat d'hora:** Un valor booleà el qual s'estableix a cert per a les *timex* que fan referència a expressions amb una precisió

¹⁰Més detalls en l'apèndix A.

d'hores (p. e.: *morning, night*). Açò es útil per a distingir hores de dates.

- **Indicador de repetició:** Un valor booleà activat quan la *timex* comença per *each* o *every* (p. e.: *each Tuesday*), o quan coincideix amb *TUnit+ly* (p. e.: *monthly*) o *Weekday+s* (p. e.: *Tuesdays*). Açò ajuda en la classificació de *timex* del tipus repetició.

En resum, la classificació de *timex* es du a terme amb una aproximació basada en aprenentatge automàtic, en la qual algunes de les característiques s'obtenen mitjançant un recurs basat en coneixement (TIMEK).

Normalització de *timex*

Com en la classificació, les característiques s'obtenen en el nivell d'element (nivell de *timex*). En la nostra proposta, el procés de normalització consisteix en dos passos: (1) l'obtenció del tipus de normalització (seguint una aproximació basada en aprenentatge automàtic), i (2) aplicació de les corresponents regles de normalització definides en el mòdul TIMEK (seguint una aproximació basada en regles) –vegeu l'apèndix A.

El primer pas assigna a cada *timex* un tipus de normalització usant les mateixes característiques s'empren per a la classificació, i el segon aplica les regles de normalització corresponents al tipus de normalització obtingut. Per a cada tipus de *timex*, definim els següents tipus de normalització i les seues regles de normalització associades:

- Per a *dates* i *hores*:
 - **ISO_explicit:** Aquest representa dates o hores explícites com ara *Feb 2010*. S'han definit regles per a convertir qualsevol format de data o hora explícita al format ISO 8601. L'exemple (47) mostra alguns casos.
 - (47) a. June 8th, 1983 → 1983-06-08
 - b. October 1999 → 1999-10
 - c. nineteen ninety → 1990
 - d. 15/03/1999 6 p.m. → 1999-03-15T18:00
 - **ISO_function:** Aquest representa dates o hores relatives com *yesterday morning*. Aquest és el tipus més complicat donat que inclou les *timex* que són relatives a una data de referència o a una

altra *timex*. El sistema aplica diferents funcions per a obtenir un valor ISO 8601 vàlid en la granularitat apropiada a partir del DCT¹¹ o una altra referència. Ací, s'utilitza la detecció dels indicadors de direcció temporal com *next* o *previous*, així com també els temps i aspectes verbals. Algunes *timex* d'aquest tipus es mostren en l'exemple (48).

(48) Document creation time (DCT): 2010-06-09

- a. 3 days ago (*pattern: NUM TUNIT ago*): DCT-3days (2010-06-06)
- b. Yesterday (*pattern: yesterday*): DCT-1day (2010-06-08)
- c. Last Friday (*pattern: last WEEKDAY*): DCT closest past Friday (2010-06-04)
- d. He will come on Friday. (*pattern: WEEKDAY*): DCT closest future Friday (2010-06-11)
- d. And the day after, we will visit Madrid. (*pattern: the TUNIT after*): previous.time.ref.+1 (2010-06-12)
- e. In October 9th, 1999, he said *three days ago*, I was living a dream. (*pattern : NUM TUNIT ago*): 1999/10/09 - 3days (1999-10-6)

Actualment, TIPSem normalitza totes les expressions relatives com a relatives a la DCT. Aleshores, falla al normalitzar aquelles que són relatives a altres *timex*, com els casos (48d) i (48e). El tractament d'aquests casos va més enllà del nivell semàntic (*i. e.*, discurs i pragmàtica).

- **Present_ref, past_ref and future_ref**: Aquests tipus són ja valors ISO 8601 per a *timex* difuses com *currently*, *some years ago*, o *in the future*. Exemples d'aquestes expressions es mostren en (49).

- (49) a. currently → present_ref
b. some years ago → past_ref

- Per a *duracions*:

- **Period**: Aquest representa duracions com *two years* o *few minutes*. S'han definit regles per a convertir períodes TIMEX3 en períodes normalitzats. L'exemple (50) mostra diferents casos. Aquestes regles són recursives i poden resoldre expressions complexes com (50e).

¹¹Del anglès *Date Creation Time*

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

- (50) a. three days → P3D
b. 50 years → P50Y
c. ten months → P10M
d. 10 minutes → PT10M
e. a year, 7 days and 5 hours → P1Y7DT5H

- Per a *repetitions*:

- **ISO_set**: Aquest representa patrons recurrents en el temps, com ara *weekly*. Hem definit regles per a obtenir una ISO 8601 vàlida per a repeticions. Les repeticions s’han de representar com a dates en les quals els valors que poden variar se substitueixen per *X*. D’aquesta forma, l’expressió (51a) es normalitza com a *1999-XX*, que significa que és vàlida per a *1999-01, 1999-02, ..., 1999-12*. Es mostren més exemples en (51).

- (51) a. every month of 1999 → 1999-XX
b. each month → XXXX-XX
c. daily → XXXX-XX-XX
d. each Thursday → XXXX-WXX-4
e. Tuesday nights → XXXX-WXX-2TNI

3.2.2 Processament d’esdeveniments

Un esdeveniment es defineix en TimeML com l’expressió lingüística de quelcom que ocorre o és verdader duran un període de temps. Com s’ha dit en la secció 2.2, els esdeveniments TimeML són classificats en les classes següents: *occurrence, state, reporting, perception, aspectual, i-action i i-state*.

El processament computacional dels esdeveniments en el llenguatge natural consisteix en dos subtasques: (i) reconeixement dels esdeveniments en el text, i (ii) la classificació d’aquestes seguint les classes TimeML. L’exemple (52) il·lustra un text pla d’entrada i l’eixida esperada després d’aquest procés.

- (52) **Input:** The TV news announced an accident in the mountains. The news reporter said that John was alive. The rescue started two hours ago.

Output:

```
The TV news <event class="reporting">announced</event> an
<event class="occurrence">accident</event> in the mountains.
The news reporter <event class="reporting">said</event> that
John was <event class="state">alive</event>.
The <event class="occurrence">rescue<event>
<event class="aspectual">started</event> two hours ago.
```

La figura 3.4 resumeix la part de l'arquitectura del TIPSem que s'encarrega del processament d'esdeveniments.

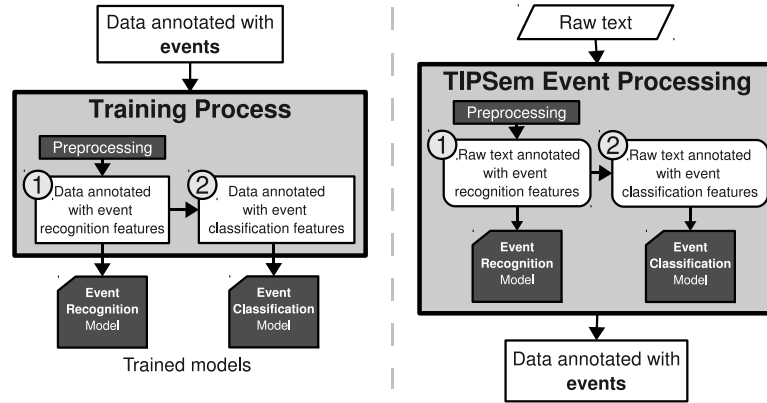


Figura 3.4: Arquitectura de processament d'esdeveniments del TIPSem

Com mostra la figura, el processament d'esdeveniments és dut a terme per una aproximació purament basada en aprenentatge automàtic sense cap regla manual ni coneixement extern.

Reconeixement d'esdeveniments

El reconeixement d'esdeveniments implica la delimitació d'aquests en el text. El problema del reconeixement d'esdeveniments es defineix en TIPSem de manera anàloga al reconeixement de *timex*. Hem considerat el cas com un problema d'etiquetatge de seqüències, com s'il·lustra en l'exemple (53).

| | | |
|------|-------------------|-----------------|
| (53) | <u>input text</u> | <u>solution</u> |
| | He | O |
| | said | B-event |
| | she | O |
| | likes | B-event |
| | him | O |
| | . | O |

Les característiques que proposem per a resoldre aquesta tasca s'obtenen en el nivell de paraula, és a dir, cada paraula té el seu conjunt de característiques.

Definim el següent conjunt de **característiques morfosintàctiques** per a abordar el problema:

- **Morfològiques:** S'empren el lema i el PoS en un context de finestra estàtica de 3 (-1,+1).

El lema és útil donat que alguns lemes és més probable que representen esdeveniments com *war*, *meeting*, o *said*; mentre que altres com *house* o *Mary* no. A l'exemple (54), les dues primeres frases són similars en el nivell morfosintàctic. L'única diferència es l'últim lema (*house* vs. *meeting*). Una aproximació basada en aprenentatge automàtic pot aprendre dels exemples que *meeting* es un EVENT, mentre que *house* no.

Pel que fa al PoS, els esdeveniments són expressats per verbs generalment (principalment verbs conjugats), nominalitzacions, adjectius, frases predicatives, o sintagmes preposicionals. Per tant, el PoS és útil per a aquesta tasca. Però, el PoS aïllat no és suficientment discriminadori per a reconèixer alguns esdeveniments. Com es mostra en l'exemple (54), no tots els verbs denoten esdeveniments segons la definició de TimeML per a l'anglès (p. e.: *to be*), ni tampoc tots els noms (p. e.: *house*).

Pel que fa al context, aquest és útil en els casos on el lema i el PoS, aïllats, no són suficients. Una finestra estàtica de 3 és el context òptim per a aquesta tasca (segons els resultats de Llorens et al. (2011a)). Per exemple, el lema *control* apareix en (54c) i (54d), però només en el segon cas representa un esdeveniment. Aquesta ambigüitat es pot solucionar consultant el context. En (54c) *control* és un modificador de *tower*, mentre que en (54d) *control* és l'estat actual de la situació.

- (54) a. He went to the house.
 b. He went to the [meeting *EVENT*].
 c. They are [working *EVENT*] at the control tower.
 d. The situation is [controlled *EVENT*].
 e. A meeting is an assembly of people for a purpose.
 f. Bruce Springsteen [sang *EVENT*] *war* in his last concert.

- **Sintàctiques:** L'estructura sintàctica ajuda en el reconeixement d'esdeveniments. L'exemple (55) mostra algunes frases en les quals el reconeixement d'esdeveniments pot beneficiar-se de considerar la informació sintàctica.

- (55) a. (S (NP America) (VP [has *EVENT*] (NP the tanks)))).
 b. (S (NP America) (VP has (VP been (VP sending (NP troops))))).

En les frases, *has* és un esdeveniment o només un verb auxiliar, dependent de la jerarquia de sintagmes verbals.

Com en el reconeixement de *timex*, la informació morfosintàctica no es suficientment discriminatòria per a reconèixer alguns esdeveniments com s'ha explicat en el capítol 2. Per aquesta raó, hi incloem les següents **característiques semàntiques**:

- **Semàntica lèxica**: La majoria dels verbs d'un text denoten esdeveniments. No obstant això, només alguns noms en condicions semàntiques específiques denoten esdeveniments. Les xarxes semàntiques aporten informació útil per a reconèixer esdeveniments nominals. Els conceptes *event* i *state* estan entre els quatre hiperònims més generals de la majoria dels esdeveniments nominals (exclouen el concepte arrel de WordNet *entity*). Per tant, la nostra aproximació emprà com a característica *top4hypers* (vegeu l'exemple 56).

(56) *Ritter led his team in a 10-hour tour.*

The *top4hypers* value for *tour* is:

tour: act => **event** => psychological_feature => abstraction

Com que molts dels esdeveniments contenen paraules que obtenen els mateixos valors per a *top4hypers*, aquesta característica incrementarà la probabilitat de representar esdeveniments per a paraules que obtinguen aquests valors, cosa que és útil per a generalitzar i poder reconèixer esdeveniments que no apareguen en les dades d'entrenament.

- **Rols semàntics**: La relació entre els rols semàntics i els esdeveniments és doble: (i) molts esdeveniments són verbs principals que tenen arguments amb diferents rols, i (ii) també molts esdeveniments poden aparèixer en diferents rols de verbs específics. Per això, considerem les següents característiques basades en rols per a reconèixer esdeveniments:

- *role (aïllat)* representa el rol que cada paraula té en un predicat. A més dels verbs principals (aquells que governen els arguments dels predicats) que normalment representen esdeveniments, aquesta característica aprèn la probabilitat dels esdeveniments d'aparèixer sota diferents rols. En (57) podem veure tres frases etiquetades amb rols amb els esdeveniments subratllats.

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

- (57) a. [The tax rates A_1] were [lower A_2].
b. [The company A_0] completed [the transaction A_1].
c. [It A_0] [might $AM-MOD$] drop out [the effort against Iraq A_4].

L'exemple (57a) mostra un esdeveniment (estat) *lower* que apareix sota el rol A2. En (57b), el rol A1 del verb *to complete* (tasca o acció a acabar) governa una ocurrència nominal de l'esdeveniment *transaction*. En (57c), només el verb principal representa un esdeveniment i cap dels seus arguments conté esdeveniments (A0, AM-MOD, A4).

En l'exemple, i en general, els esdeveniments apareixen com a verbs principals quan són verbals, i és més probable que apareguen sota la influència dels rols A1 o A2 quan són nominals o adjectivals. Per exemple, és menys probable trobar esdeveniments sota rols com A0, AM-MOD o A4.

- *role + token* és la combinació de la paraula i el rol semàntic que té.

- (58) a. [He A_0] made [an offer A_1].
b. [He A_0] made [a cake A_1].
c. [He A_0] completed [the transaction A_1].
d. [He A_0] completed [the puzzle A_1].

En les frases de l'exemple (58a) i (58b), *offer* i *cake* són noms que tenen el rol A1. No obstant això, només *offer* representa un esdeveniment. Anàlogament, en les frases (58c) i (58d), *transaction* i *puzzle* tenen l'A1 però només *transaction* és un esdeveniment. Això és perquè no totes les paraules representen esdeveniments, sinó només aquelles que signifiquen accions, processos, o estats. *role + token* mescla la paraula i el rol que té per a capturar aquesta informació.

- *role + verb* és la combinació del rol i el verb que governa el predicat. Aquesta característica diferencia rols que depenen de verbs diferents. Això és particularment important en el cas dels rols numerats (A0, A1, etc.), els quals poden significar coses diferents quan depenen de verbs diferents.

- (59) a. [The tax rates A_1] were [lower A_2].
b. [The tax rates A_0] raised [10% A_2] [to \$11 A_4] [from 10 A_3].
c. [He A_0] carried out [an experiment A_1].
d. [He A_0] ate [an apple A_1].

En les frases (59a) i (59b) es pot veure que el rol A2 del verb *to be* és més probable que continga un esdeveniment (estat) que el

- rol A2 del verb *to rise*. De la mateixa manera, en (59c) i (59d), el rol A1 de *to carry out* és més probable que continga una acció o procés que el rol A1 de *to eat*, el qual representa la cosa menjada.
- *role + top4hypers* és la combinació del rol i els quatre hiperònims més generals de la paraula, exceptuant *entity*. Aquesta característica generalitza *role + token*, i obté el mateix valor per a totes les paraules que compartiquen els quatre hiperònims més generals i que apareguen sota el mateix rol.

(60) [The killing _{A0}] took place [this morning _{AM-TMP}].

En l'exemple (60), *killing* obté el valor (*A0+action => event => psychological_feature => abstraction*) per a *role+top4hypers*. Aquest valor serà compartit per lemes d'esdeveniments semànticament similars que no apareguen en l'entrenament (p. e.: murder, execution, homicide, etc.). Aquesta característica ajuda el model a generalitzar quins *top4hypers* solen representar esdeveniments quan apareixen sota rols específics (p. e.: A0).

La figura 3.5 il·lustra les característiques descrites.

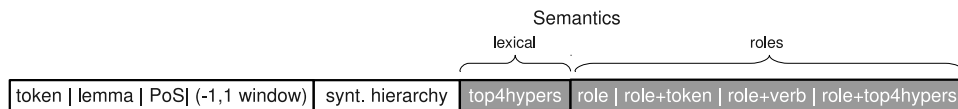


Figura 3.5: Característiques del TIPSem per al reconeixement d'esdeveniments

Considerant la frase (61), l'exemple (62) mostra els valors d'aquest vector obtinguts per la paraula *war*.

(61) The phony war has finished and the referendum campaign has begun.

(62) phony | phony | JJ
war | war | NN [NP|act > event > ...|A1|A1+war|A1+finish|A1+act > event > ...
has | have | AUX

Classificació d'esdeveniments

Com es descriu a la secció 2.2, hi ha set classes d'esdeveniments en TimeML:

- *Reporting*. Acció de narrar un esdeveniment (p. e.: *say, report*).
- *Perception*. Percepció física d'un esdeveniment (p. e.: *said, hear*).
- *Aspectual*. Predicació aspectual d'un esdeveniment (p. e.: *start, continue*).
- *I_Action*. Acció intencional (p. e.: *try, attempt*).
- *I_State*. Estat intencional (p. e.: *feel, hope*).
- *State*. Circumstància verdadera per un període de temps (p. e.: *rainy, bad*).
- *Occurrence*. Esdeveniments generals (p. e.: *came, conference*).

Els esdeveniments reconeguts pel model de reconeixement són classificats com es mostra en l'exemple (63).

| | | |
|------|--------------|-----------------|
| (63) | <u>event</u> | <u>solution</u> |
| | went | occurrence |
| | said | reporting |
| | meeting | occurrence |
| | started | aspectual |
| | tried | i.action |

La mateixa estratègia usada per a les *timex* ha sigut emprada per a convertir les característiques basades en paraules del reconeixement d'esdeveniments en característiques basades en elements per a la classificació. A més, les mateixes característiques emprades per al reconeixement han sigut emprades per a la classificació.

Les característiques semàntiques poden contribuir a la classificació d'esdeveniments en els aspectes següents:

- **Semàntica lèxica:** Les xarxes semàntiques podrien afectar la classificació d'esdeveniments no vistos en l'entrenament generalitzant les classes verbals. Analitzant l'exemple (64) i considerant que *announced* fóra una paraula no continguda en l'entrenament, la informació semàntica podria ajudar a classificar-la correctament com a *reporting*, ja que altres esdeveniments d'aquesta classe com *say, tell* o *ask*, estan relacionats amb la classe semàntica *communicate*.

- (64) President Bush [announced *REPORTING*] the dispatching of U.S. troops.
 Tokens: announced
 PoS: VBD Semantic role: V
 4 top semantic class: announce => inform => communicate => act

- **Rols semàntics:** El rol dels arguments d'un verb en particular poden ser útils en la classificació d'esdeveniments ambigus, perquè un lexema pot representar diferents classes d'esdeveniments quan té el mateix rol de diferents verbs. Hem seleccionat un exemple del corpus del TempEval-2 (65).

- (65) [They _{A0}] said [[they _{A0}] reached [an agreement_{I.STATE} for the acquisition _{A1}] _{A1}].
 [The offer _{A0}] violates [a standstill agreement_{OCCURRENCE} _{A1}].

En la primera frase, *agreement* té el rol A1 governat pel verb *reach* i governa l'esdeveniment *acquisition*. Això podria ajudar a determinar que és un estat intencional. En la segona frase, *agreement* té el rol A1 governat pel verb *violate*. Això podria indicar que hi ha un acord real que pot ser violat, i, per tant, és un esdeveniment de classe ocurrència.

3.2.3 Processament de relacions temporals

El processament de relacions temporals es divideix en dues tasques principals: la identificació i la categorització.

La **identificació de relacions temporals** implica la selecció de parells d'entitats (event-event, event-*timex*, o *timex-timex*) relacionats, i el tipus d'enllaç (TLINK, SLINK, or ALINK) que representen. Les relacions amb *timex* o entre esdeveniments principals són TLINK. Les relacions entre esdeveniments subordinats són SLINK. Finalment, les relacions entre un esdeveniment aspectual i l'esdeveniment que modifica són ALINK.

TIPSem inclou la següent estratègia basada en regles per a identificar relacions temporals:

1. Si un predicat té més d'un esdeveniment, aleshores es relaciona l'esdeveniment principal amb: (i) el corresponent esdeveniment aspectual (ALINK) si la classe es aspectual, o (ii) l'esdeveniment subordinat (SLINK) en els altres casos.

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

2. Si una frase té una o més *timex*, aleshores relaciona l'esdeveniment sintàcticament més proper amb cada una (TLINKs) – Totes les *timex* han d'estar relacionades com a mínim amb un esdeveniment.
3. Relacionar els esdeveniments principals (main-events¹²) de les frases consecutives (TLINKs).

Encara que TIPSem inclou aquesta estratègia d'identificació de relacions temporals, la tasca central analitzada en aquesta tesi és la categorització de relacions temporals.

La **categorització de relacions temporals** implica assignar una categoria a cada relació identificada de la taxonomia definida per Allen (1983) (vegeu la secció 2.1). Aquesta tasca obté un ordre específic en el temps dels esdeveniments d'un text i serveix com a entrada de sistemes més complexos que requereixen capacitats de raonament temporal.

En els exercicis d'avaluació internacional TempEval-1 (Verhagen et al., 2007) i TempEval-2 (Verhagen et al., 2010), la tasca de processament de relacions temporals s'ha centrat en la categorització –la part d'identificació no ha sigut avaluada. A causa de la complexitat de les relacions TimeML, les avaluacions TempEval han simplificat la tasca reduint el nombre de categories. En aquesta tesi, seguim la simplificació per a poder comparar els resultats amb els obtinguts per altres investigadors en els treballs relacionats.

En el exercici d'avaluació TempEval-2, el problema del processament de les relacions temporals es va dividir en quatre subtasques: *event-timex*, *event-dct*, *main-events* i *subordinated-events*. A més, les categories es van reduir a sis: les tres relacions bàsiques (*before*, *after*, i *overlap*), dues menys específiques per a casos ambigus (*before-or-overlap* i *overlap-or-after*), i finalment *vague* per a aquelles relacions on no es podia establir una relació concreta.

Donat un corpus textual en llenguatge natural en el qual les *timex* i els esdeveniments estaven ja anotats, aquestes subtasques es defineixen com:

- **event-timex**: Aquesta consisteix en categoritzar la relació temporal entre un esdeveniment i una *timex* en la mateixa frase (només si l'esdeveniment domina sintàcticament la *timex* o l'esdeveniment i la *timex* estan en el mateix sintagma nominal). Aquesta tasca va ser anomenada *Task C* en TempEval-2. L'exemple (66) il·lustra el text d'entrada i d'eixida després del processament descrit.

¹²Els més alts en la jerarquia sintàctica.

(66) He was <EVENT eid=1>born</EVENT> in <TIMEX3 tid=1>1999</TIMEX3>.

Input: <TLINK lid=1.11"reltype="NONE1.leid="11.1tid="1"/>
 Output: <TLINK lid=1.11"reltype="OVERLAP1.leid="11.1tid="1"/>

Aquesta categorització indica que *born* se solapa amb *1999*.

- **event-dct**: Aquesta consisteix en la categorització de les relacions temporals entre un esdeveniment i la data de creació del document (DCT). La tasca va ser anomenada *Task D* en TempEval-2. Un exemple es mostra en (67).

(67) DCT: 2001-01-15 (tid="0")

He was <EVENT eid=1>born</EVENT> in <TIMEX3 tid=1>1999</TIMEX3>.

Input: <TLINK lid=1.11"reltype="NONE1.leid="11.1tid="0"/>
 Output: <TLINK lid=1.11"reltype="BEFORE1.leid="11.1tid="0"/>

Aquesta categorització indica que *born* (1999) és abans que la data en la qual es va crear o escriure el document (2001-01-15).

- **main_event-event (intersentential)**: Aquesta consisteix en la categorització de les relacions temporals entre dos esdeveniments principals de frases consecutives. La tasca correspon a la *Task E* en TempEval-2 –se'n mostra un exemple en (68).

(68) He was <EVENT eid="1">born</EVENT> in London.
 He <EVENT eid="2">obtained</EVENT> his degree
 in computer science at Cambridge University.
 He had <EVENT eid="3">studied</EVENT> previously in Oxford.

Input:
 <TLINK lid=1.11"reltype="NONE1.leid="11.leid="2"/>
 <TLINK lid=1.12"reltype="NONE1.leid="21.leid="3"/>

Output:
 <TLINK lid=1.11"reltype="BEFORE1.leid="11.leid="2"/>
 <TLINK lid=1.12"reltype="AFTER1.leid="21.leid="3"/>

Aquesta categorització vol dir que *born* va ser abans que *obtained* i *obtained* després que *studied*.

- **sub_event-event (intrasentential)**: Aquesta determina la relació temporal entre dos esdeveniments, on un domina sintàcticament l'altre. La tasca es va anomenar *Task F* al TempEval-2. S'en mostra un exemple en (69).

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

(69) We <EVENT eid=1>saw</EVENT> an <EVENT eid=2>explosion</EVENT>.

Input: <TLINK lid=1·11"reltype="NONE1·leid="11·leid="2"/>
Output: <TLINK lid=1·11"reltype="OVERLAP1·leid="11·leid="2"/>

Aquesta categorització com a *overlap* vol dir que *saw* i *explosion* se superposen en el temps.

La figura 3.6 resumeix l'arquitectura de TIPSem per a la part de processament de relacions temporals.

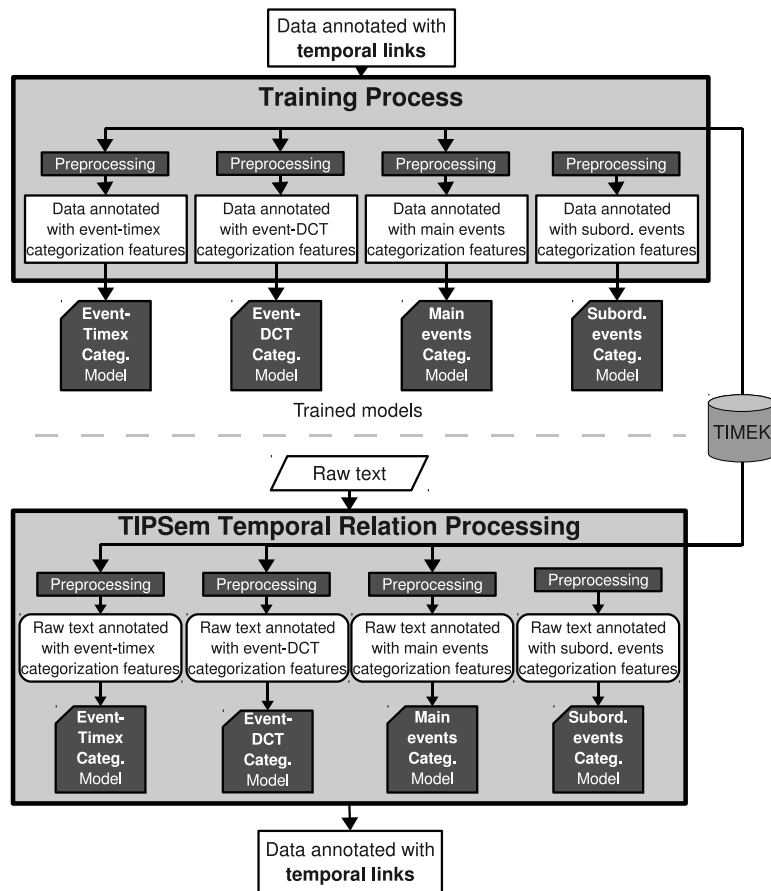


Figura 3.6: Arquitectura de processament de relacions temporals de TIPSem

Com es mostra a la figura, el processament de relacions temporals és dut a terme per models apresos automàticament. Tres d'aquests utilitzen una

base de coneixements externa (TIMEK) per a obtenir característiques. Cada una de les tasques descrites té les seues pròpies característiques. Aquestes es detallen en les subseccions següents.

Relacions event-*timex*

TIPSem utilitza les següents característiques per a conèixer un model de classificació de les relacions event-*timex*.

Les **característiques morfosintàctiques** són:

- *Preposició governant (Heading PP) de l'esdeveniment*. Si l'esdeveniment està contingut en un sintagma preposicional (p. e.: *before the meeting*), el valor d'aquesta característica és la preposició governant (p. e.: *before*); si no ho està, el valor es deixa com a -. Preposicions com ara *after* o *in* són assenyaladores de relacions, i freqüentment assenyalen la categoria de la relació.
- *Preposició governant (Heading PP) de la timex*. Si la *timex* està continguda en un sintagma preposicional (p. e.: *after 1999*), el valor d'aquesta característica és la preposició governant (p. e.: *after*); si no ho està, el valor es deixa com a -. Preposicions com ara *before*, *from*, *to*, etc., són assenyaladores de relacions (TimeML SIGNAL), i aquesta característica captura aquesta propietat de les *timex*.
- *Interval*. Aquesta característica booleana es considera falsa, si no apareix alguna paraula que indique interval (*i. e.*, -, /, *from*, *to*) prop de la *timex* de la relació. Això es útil per a identificar les categories overlap-and-after i overlap-and-before.

Les **característiques semàntiques** són:

- **Rols semàntics**
 - *Nucli de subordinació amb rol temporal (TRoleSBARHead) de l'esdeveniment*. Si l'esdeveniment està contingut en una subordinació amb rol temporal, el valor d'aquesta característica és la conjunció que governa la subordinació, en cas contrari, -. Aquesta propietat del rol temporal pot ser útil per a categoritzar la relació temporal. Conjuncions com *when*, *after*, i *while* denoten TLINKs i aporten informació sobre la seua categoria. Si el nucli és *when* o *while* la categoria de l'enllaç és més probable que

sigui *overlap*, mentre que si és *after* o *once* la categoria és més probable que sigui *after* – vegeu l'exemple (70). No obstant això, la característica no és suficientment discriminatòria per a categoritzar totes les relacions. Aquestes conjuncions, en aparèixer en altres papers semàntics no poden estar relacionats amb la informació temporal (70c). A més, tenint en compte els elements continguts en els rols semàntics temporals, evita haver de definir una llista *ad hoc* d'assenyaladors temporals per a cada idioma.

- (70) a. [After the Iraqi annexation of Kuwait $AM-TMP$],
oil prices rose.
b. He was rescued, [while the thieves were stealing
the safe $AM-TMP$].
c. That shop is [next to the bank we passed $AM-LOC$].

- *Nucli de subordinació amb rol temporal (TRoleSBARHead) de la timex*. Si la *timex* està continguda en un rol semàntic temporal de subordinació, la conjunció nucli de la subordinació és el valor d'aquesta característica, en cas contrari, -. És equivalent a la característica anterior, però aplicada a l'element *timex*.

• Semàntica temporal

- *Posició temporal (valdiff)*. De vegades una frase conté diverses *timex* i esdeveniments. Si l'esdeveniment de la relació no està sintàcticament vinculat a la *timex* de la relació, sinó a una altra *timex* (*timex'*), la característica de la posició temporal (*valdiff*) representa si aquesta altra *timex* és *before* (abans), *equal* (igual), o *after* (després) que la *timex* de la relació (la diferència de valor). En cas contrari, el valor s'estableix en *equal* per defecte. Com es mostra en (71), si el sistema ha de classificar la relació e1-t2, val la pena tenir en compte que e1 està sintàcticament relacionat amb t1, i que t1 (1999) és major que t2 (1998), de manera que el *valdiff* és *after*. No obstant això, si el sistema ha de categoritzar e1-t1, que estan sintàcticament relacionades, la classificació no depèn d'altres *timex* (*default_valdiff* = *equal*).

(71) John was born_{e1} in 1999_{t1} and Sylvia was born_{e2} in 1998_{t2}.

```
valdiff (e1,t1): equal
valdiff (e1,t2): after
valdiff (e2,t1): before
valdiff (e2,t2): equal
```

La figura 3.7 il·lustra el vector de característiques usat en la categorització event-*timex*.

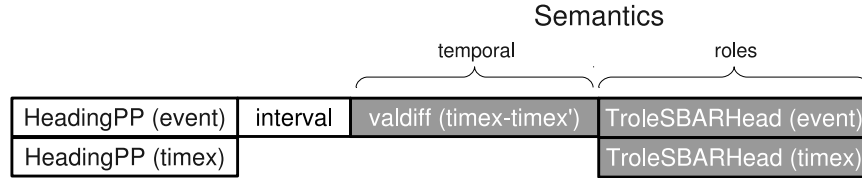


Figura 3.7: Característiques de TIPSem per a event-*timex*

En (72) apareix, en un cas simple, un exemple de les característiques descrites.

```
(72) He <event eid=e1>lived</event> here from <timex tid=t1>1903</timex>.
      categorize (e1,t1)
```

```
HeadingPP (event): -
HeadingPP (timex): from
interval: true
valdiff: equal
TRoleSBARHead (event): -
TRoleSBARHead (timex): -
```

En (72), la relació entre *lived* i *1903* representa una relació overlap-and-after perquè *lived* comença el *1903* però continua fins a un punt indefinit del futur i per tant l'esdeveniment també ocorre després de *1903*. En aquest cas, les característiques *interval* i *HeadingPP(timex)* són rellevants.

Relacions event-DCT

Per a determinar la relació entre un esdeveniment i la data de creació del document (event-DCT), TIPSem usa les característiques definides per a les relacions event-*timex*, excepte *interval*, que no s'utilitza. D'altra banda, s'hi afegeixen dues característiques morfosintàctiques noves.

Les **característiques morfosintàctiques** per a aquesta tasca són les següents:

- *HeadingPP (event)* i *HeadingPP (timex)*. Aquesta ja va ser definida per a les relacions event-*timex*. No obstant això, en aquest cas, el nucli del sintagma de la *timex* es calcula respecte a la *timex* sintàcticament més propera a l'esdeveniment de la relació (si existeix). En cas contrari el valor es deixa a -.

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

- *Syntactic relation*. Si l'esdeveniment de la relació i la *timex* més propera estan en la mateixa frase, el valor és *sentence (frase)*. Si estan en la mateixa clàusula¹³ el valor és *sub-sent*. En cas contrari, el valor és *sentence*.
- *Temps i aspecte verbal*. Fa referència al temps i l'aspecte dels esdeveniments verbals, o al temps i l'aspecte dels verbs que regeixen esdeveniments no verbals. Aquesta característica és rellevant, perquè el DCT coincideix amb el moment de la parla (Reichenbach, 1947), i el temps i aspecte ofereixen informació sobre la ubicació relativa al DCT dels esdeveniments.

Les **característiques semàntiques** afegides són:

- **Semàntica temporal**
 - *Posició temporal (valdiff)*. En aquest cas, *valdiff* es calcula respecte a la DCT. Si l'esdeveniment està sintàcticament vinculat amb una *timex*, la característica de la posició de temps indica si aquesta *timex* és *before*, *equal* o *after* de la DCT (la diferència de valor). En cas contrari, el valor s'estableix en *equal* per defecte.

La figura 3.8 il·lustra el vector de característiques emprat en la categorització de relacions event-DCT.

| Semantics | | | | |
|-------------------|----------------------|------------------------|---------------------|-----------------------|
| | | | temporal | roles |
| HeadingPP (event) | tense-aspect (event) | synt-rel (event-timex) | valdiff (timex-DCT) | TroleSBARHead (event) |
| HeadingPP (timex) | | | | TroleSBARHead (timex) |

Figura 3.8: Característiques de TIPSem per a event-DCT

En (73) es mostra un exemple d'aquestes característiques.

(73) Document creation time: 2010-06-09 (t0)

```

He <event eid=e1>lived</event> here in <timex tid=t1>1999</timex>.
categorize (e1,t0)
HeadingPP (event): -
HeadingPP (timex): in
valdiff: before

```

¹³Frase simple o oració subordinada d'una oració complexa.

Tense-aspect: past-simple
TRoleSBARHead (event): -
TRoleSBARHead (timex): -

L'exemple (73) mostra un esdeveniment *lived* que té lloc abans (*before*) que la DCT. Les característiques ajuden perquè *lived* està en temps passat i perquè es relaciona amb una *timex* 1999 que és prèvia a la DCT 2010.

Relacions d'esdeveniments principals (Main Event)

Les relacions entre els dos esdeveniments principals (event1 i event2) són diferents dels dos casos anteriors, perquè hi ha dos esdeveniments implicats en lloc d'un. En aquest cas, les característiques del TIPSem se centren en aquests esdeveniments i les *timex* sintàcticament més properes que pot haver-hi, respectivament.

Les **característiques morfosintàctiques** són:

- *Temps i aspecte (event1)*. El temps i l'aspecte dels esdeveniments verbals, o el temps i l'aspecte del verbs que regeixen esdeveniments no verbals.
- *Temps i aspecte (event2)*. El temps i l'aspecte per al segon esdeveniment principal (Event2).
- *Temps i aspecte (event1-event2)*. Una combinació de les dues característiques anteriors. Els esdeveniments normalment segueixen un ordre que avança en el temps, que correspon al ordre en què apareixen en el text. Aquesta característica ajuda a trobar la informació de temps i aspecte verbal en combinació (p. e.: *past/simple-past/perfect*), que podria apuntar al fet que s'ha trencat l'ordre regular de precedència d'esdeveniments.
- *Mateixa frase (same sentence) (event1-event2)*. Aquesta característica indica si els dos esdeveniments estan en la mateixa frase o en frases diferents. Normalment, les relacions entre esdeveniments principals estan en frases consecutives. No obstant això, de vegades dos esdeveniments principals apareixen en la mateixa frase.

Les **característiques semàntiques** són:

- **Semàntica temporal**

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

- *Posició temporal (valdiff)*. En aquest cas, *valdiff* es calcula respecte a dues *timex* relacionades amb cada un dels esdeveniments de la relació. Si ambdós esdeveniments estan sintàcticament relacionats amb diferents *timex*, la posició temporal representa si la *timex* relacionada amb el primer esdeveniment (*event1_timex*) és *before*, *equal* o *after* que la *timex* relacionada amb el segon esdeveniment (*event2_timex*). En cas contrari, el valor és *equal* per defecte.

Els rols semàntics no s'utilitzen en aquest cas perquè normalment afecten relacions temporals intraoracionals i el seu ús podria ser perjudicial per al model.

La figura 3.9 il·lustra el vector de característiques emprat en la categorització de relacions entre esdeveniments principals.

| | | | |
|-----------------------|------------------------------|---------------------------|---------------------------------------|
| tense-aspect (event1) | tense-aspect (event1-event2) | same-sent (event1-event2) | valdiff (event1_timex – event2_timex) |
| tense-aspect (event2) | | | |

Semantics
temporal

Figura 3.9: Característiques del TIPSem per a relacions *main-event*

En (74) es mostra un exemple de les característiques descrites.

(74) He <event eid=e1>came</event> that night.
 She had already <event eid=e2>gone</event>.
 categorize (e1,e2)
 tense-aspect (event1): past_simple
 tense-aspect (event2): past_perfect
 tense-aspect (event1-event2): past_simple - past_perfect
 same-sent: false
 valdiff: equal

L'exemple (74) il·lustra una relació *after* entre l'esdeveniment *came* i l'esdeveniment *gone*. En aquest tipus de relacions, el temps i l'aspecte verbals tenen un paper cabdal perquè mostren que el primer esdeveniment passa després que el segon. La característica *valdiff* es deixa amb el valor per defecte, perquè només el primer esdeveniment té una *timex* associada (i. e., *that night*). Si el segon esdeveniment estiguera relacionat també amb una *timex* (p. e.: *the day before*) la característica *valdiff* ajudaria en la categorització.

Relacions entre esdeveniment subordinats (Subordinated Event)

Les relacions entre dos esdeveniments subordinats (*event1* i *event2*) són diferents del cas anterior, ja que hi ha un esdeveniment principal i un de subordinat. Un cop més, les característiques del TIPSem se centren en aquests esdeveniments i les *timex* sintàcticament més properes que pot haver-hi, respectivament.

Per a la categorització dels esdeveniments subordinats, TIPSem utilitza el nucli dels esdeveniments subordinats en rols temporals (si és el cas), la preposició nucli dels sintagmes preposicionals que continguen els esdeveniments (si és el cas), així com el temps i l'aspecte verbal d'ambdós esdeveniments en una característica combinada.

Les **característiques morfosintàctiques** són:

- *HeadingPP (event1)* i *HeadingPP (event2)*. Com ja s'ha definit, si els esdeveniments estan continguts en sintagmes preposicionals aquesta característica consisteix en la preposició nucli. En cas contrari, el valor es deixa a -.
- *Tense and aspect: (event1), (event2), (event1-event2)*. Definits de la mateixa manera que s'han definit per als esdeveniments principals.

Les **característiques semàntiques** són:

- **Semàntica lèxica**
 - *TimeML class (event1)*. Aquesta consisteix en el valor de l'atribut de classe per a l'*event1* (p. e.: OCCURRENCE, STATE, I_ACTION).
 - *TimeML class (event2)*. El valor de l'atribut de classe per l'*event2*.

Les classes TimeML dels esdeveniments poden ajudar en casos com (75).

```
(75) We will <event eid=1 class=I_ACTION>decide</event> where  
to <event eid="2" class=OCCURRENCE>relocate</event> this.  
<TLINK lid=l1 reltype=BEFORE leid=1 leid=2 />
```

En aquest cas, saber que *decide* és un I_ACTION i *relocate* és un OCCURRENCE ajuda a determinar que *deciding* és anterior a *relocating*. Algunes classes d'esdeveniments en relacions de subordinació determinen quines categories són més probables

3.2. ARQUITECTURA ESPECÍFICA DE CADA TASCA

(p. e.: $\text{LACTION+OCCURRENCE} \rightarrow \text{before}$). Aquesta relació semàntica no es pot fer usant només característiques relatives al temps verbal, perquè el futur de *decide* podria indicar la relació inversa. No obstant això, el temps verbal és important en alguns casos com en *We think she played very well*, en el qual la relació és *after*.

La figura 3.10 il·lustra el vector de característiques usat en la categorització d'esdeveniments subordinats.

| | | | Semantics lexical |
|--------------------|-----------------------|------------------------------|-----------------------|
| HeadingPP (event1) | tense-aspect (event1) | tense-aspect (event1-event2) | TimeML class (event1) |
| HeadingPP (event2) | tense-aspect (event2) | | TimeML class (event2) |

Figura 3.10: Característiques de TIPSem per a *subordinated-event*

En (76) es mostra un exemple de les característiques descrites.

(76) He <event eid=e1>said</event> he will <event eid=e2>come</event>.
categorize (e1,e2)

HeadingPP (event1): -

HeadingPP (event2): -

Tense-aspect (event1): past_simple

Tense-aspect (event2): future_simple

Tense-aspect (event1-event2): past_simple - future_simple

TimeML class (event1): REPORTING

TimeML class (event2): OCCURRENCE

L'exemple (76) mostra una relació de subordinació amb categoria *before* (*said* ocorre abans que *come*). En aquest cas, el temps verbal és rellevant per a la categorització.

En conclusió, TIPSem és una aproximació híbrida que està clarament dividida en tres mòduls que corresponen a diferents subtasques. L'aproxi-

mació està basada principalment en aprenentatge automàtic. La següent secció descriu les tècniques d'aprenentatge automàtic aplicades a cada part.

3.3 Selecció de tècniques d'aprenentatge automàtic

Els vectors de característiques que TIPSem defineix per a cada subtasca s'utilitzen per a entrenar els models aplicant diferents tècniques d'aprenentatge automàtic. En aquesta tesi, hem considerat dues tècniques àmpliament conegudes: camps aleatoris condicionals (CRF) i màquines de suport vectorial (SVM). Ambdues tècniques són molt populars entre els treballs relacionats i mostren un bon rendiment en les tasques proposades.

Els **CRF** (Lafferty, McCallum, & Pereira, 2001) són models gràfics no dirigits, un cas particular de màquines d'estats finits entrenades condicionalment. La figura 3.11 il·lustra el model.

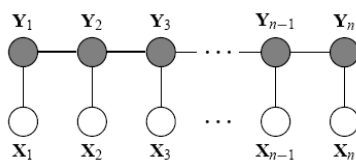


Figura 3.11: Camps Aleatoris Condicionals

Per a aplicar CRF, se suposa que X és una variable aleatòria sobre seqüències de dades en ser etiquetades, i Y és una variable aleatòria sobre les seqüències de l'etiqueta corresponent, on tots els components (Y_i) són membres d'un alfabet γ .

En les tasques de reconeixement, X representa les frases i Y les anotacions possibles d'aquestes frases, amb γ com al conjunt de possibles etiquetes IOB2¹⁴. En les tasques de classificació, X oscil·la sobre els elements i Y , sobre les possibles classificacions, normalitzacions o categoritzacions dels elements, amb γ com al conjunt de possibles classes o categories.

Les variables aleatòries X i Y es distribueixen de forma conjunta, i els CRF construeixen un model condicional a partir de l'observació emparellada de les seqüències d'etiquetes: $p(Y|X)$. Gràficament, CRF estan representats pels grafs no dirigits, $G = (V, E)$ talque $I = (Y_v)$, $v \in V$, de manera que Y és indexada pels vèrtexs de G . Llavors (X, Y) és un *camp aleatori condicional*

¹⁴Format IOB2: (B) begin, (I) inside, i (O) out.

3.3. SELECCIÓ DE TÈCNiques D'APRENENTATGE AUTOMÀTIC

si les variables Y_v obeeixen la propietat de Markov respecte al graf quan està condicionat per X :

$$P(Y_v|X, Y_w, v \neq w) = P(Y_v|X, Y_w, v \sim w),$$

on $v \sim w$ significa que Y_v i Y_w són veïns connectats en G .

CRF és una popular i eficaç tècnica d'aprenentatge supervisat per als problemes d'etiquetatge seqüencial. CRF s'han aplicat amb èxit en l'anàlisi superficial (Sha & Pereira, 2003) i el reconeixement d'entitats anomenades (McCallum & Li, 2003). En concret, en treballs relacionats amb el processament d'informació temporal, els CRF s'han utilitzat per a reconèixer les *time*s (Ahn, van Rantwijk, & de Rijke, 2007; UzZaman & Allen, 2010; Kolya et al., 2010).

Els **SVM** (Cortes & Vapnik, 1995) són classificadors binaris, àmpliament utilitzats en PLN. Gràficament, els exemples d'entrenament de cada classe es representen per punts en l'espai i una SVM construeix un hiperplà en un espai n -dimensional ¹⁵, que separa els punts relacionats a una categoria dels altres amb el major marge. La figura 3.12 mostra un espai d'entrada i un hiperplà après, que separa les classes en un espai de característiques en 3 dimensions.

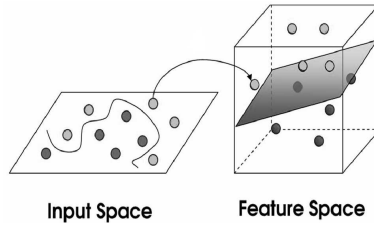


Figura 3.12: SVM

Per a aplicar SVM, es plantegen les subtasques descrites del processament de la informació temporal com problemes de classificació multiclasse. La tècnica SVM es pot estendre a aquests problemes mitjançant la traducció del problema multiclasse a múltiples problemes de classificació binària, que decideixen si un exemple pertany o no a una classe. Hi ha dues estratègies comunes per a reduir un problema multiclasse: un-contra-tots i un-contra-un. En aquesta tesi, s'utilitza la primera ¹⁶. Aquesta estratègia genera

¹⁵El nombre de dimensions depèn del nombre de característiques i el nucli utilitzat.

¹⁶Perquè ofereix un millor rendiment en les tasques abordades.

n classificadors binaris que decideixen el grau de pertinença a cada classe diferent.

En els treballs relacionats amb el processament d'informació temporal, els SVM han estat aplicats per molts autors, tant per a tasques de reconeixement com de classificació (Mani et al., 2006; Hepple et al., 2007; Bethard & Martin, 2006, 2007; Cheng et al., 2007; Min et al., 2007; March & Baldwin, 2008).

Hem integrat les següents implementacions de CRF i SVM en el sistema TIPSem.

Per a CRF, CRF++ toolkit¹⁷ ha sigut emprada, i el procés d'aprenentatge s'ha dut a terme amb els paràmetres següents: algorisme *CRF-L2* i hyper-paràmetre $C=1$.

Per a SVM, YamCha software¹⁸ ha sigut emprada, i el procés d'aprenentatge s'ha dut a terme amb els paràmetres següents: *one-vs-all*, $C = 1$ i *polynomial-degree=2*.

Les configuracions dels paràmetres descrits van ser seleccionades pel fet que van obtenir els millors resultats en l'experimentació preliminar (vegeu l'apèndix B). Aquests paràmetres ofereixen un bon equilibri entre la generalització dels models per fer front als exemples no vistos, i el sobreajust (o sobreentrenament) dels models a les dades d'entrenament.

Com que l'arquitectura TIPSem es divideix en un conjunt de mòduls, cada un pot utilitzar la tècnica d'aprenentatge automàtic que millor s'adapte a les seues necessitats. Per a la selecció de tècniques, es va portar a terme una experimentació comparativa amb les dues tècniques. La taula 3.1 mostra la tècnica que ha obtingut els millors resultats per a cada tasca (els detalls sobre els resultats es poden consultar en l'apèndix B).

Els CRF superaren els SVM en el reconeixement de *timex* i d'esdeveniments, i en la categorització de relacions temporals. Això és degut al fet que, en aquestes tasques, la classificació d'un element sovint depèn de la classificació i les característiques dels elements adjacents: la seqüència és rellevant. CRF encaixa en el reconeixement de *timex* i d'esdeveniments perquè han estat definits com a problemes d'etiquetatge de seqüències. No només la seqüència de les paraules, sinó també la morfològica, la de l'estructura sintàctica i semàntica, present en les característiques anteriorment descrites del nostre enfocament, es beneficien de l'ús d'aquesta tècnica d'aprenentatge. Per a les tasques de classificació, aquesta tècnica captura les dependències entre la seqüència de les característiques i les etiquetes. Això ha estat útil

¹⁷<http://crfpp.sourceforge.net/>

¹⁸<http://chasen.org/~taku/software/YamCha/>

3.4. EXTENSIÓ AL CASTELLÀ, ITALIÀ I XINÉS

| Module | Subtask | ML Technique |
|------------------------------|--------------------|--------------|
| Timex Processing | Recognition | CRF |
| | Classification | SVM |
| | Normalization type | SVM |
| Event Processing | Recognition | CRF |
| | Classification | SVM |
| Temporal Relation Processing | event-timex | CRF |
| | event-DCT | CRF |
| | main events | CRF |
| | subord. events | CRF |

Taula 3.1: Selecció de tècniques d'ML

en relació amb les tasques de categorització de relacions temporals, però no per a la classificació de *timex* i esdeveniments, ni en la classificació dels tipus de normalització de *timex*.

No obstant això, els SVM han obtingut millors resultats en la classificació de *timex* i esdeveniments, i en la normalització TIMEX tipus de classificació. Això és degut al fet que, en aquestes tasques, la classificació d'un element és normalment independent de la classificació dels elements adjacents: la importància de la seqüència és més baixa.

En poques paraules, prenent els resultats d'experimentació com a base empírica per a triar les tècniques d'ML, TIPSem utilitza CRF per al reconeixement i categorització, i SVM per a la classificació, incloent-hi el tipus de normalització.

3.4 TIPSem multilingüe: Extensió al castellà, italià i xinès

TIPSem va ser desenvolupat inicialment per a l'anglès. Gràcies a la seua base d'aprenentatge automàtic, la seua extensió a altres idiomes només depèn de l'existència de dades TimeML anotades i els instruments adequats de PLN per al processament previ (per exemple, els etiquetadors de PoS) en els idiomes de destinació. Només la base de coneixements TIMEK ha de ser manualment estesa a la llengua de destinació.

TIPSem s'ha estès a l'espanyol, l'italià i el xinès. En les següents subseccions es detallen les particularitats de cada extensió, els mòduls estesos, i el tipus de semàntica utilitzat.

TIPSem castellà: TIPSemES

Hem dut a terme dues adaptacions de TIPSem a l'espanyol.

En primer lloc, TIPSem es va estendre a l'espanyol per a ser avaluat en l'exercici TempEval-2 (Llorens et al., 2010b). Les dades sobre els quals el sistema havia de ser aplicat en aquest fòrum d'avaluació és un subconjunt del corpus Ancora (Taulé et al., 2008).

Ancora ja està anotat en el nivell lèxic, morfosintàctic i semàntic. Per tant, l'enfocament espanyol per a TIPSem ha obtingut els valors de les característiques de l'anotació d'Ancora.

Només els quatre primers hiperònims es van obtenir d'EuroWordNet (Vossen, 1998) i TIMEK es va estendre de manera manual amb el coneixement temporal en espanyol.

En segon lloc, després de TempEval-2, la versió en espanyol de TIPSem ha estat millorada per a treballar amb qualsevol text d'entrada ¹⁹, en lloc de fer-ho només amb el corpus Ancora. Per això, són necessàries diferents eines de la PLN per a processar l'entrada de text en el nivell morfosintàctic i semàntic.

Pel que fa al tractament morfosintàctic, TIPSemES utilitza l'eina FreeLing ²⁰ (Padró et al., 2010). FreeLing inclou segmentació d'oracions, tokenització, lematització, etiquetatge POS i l'anàlisi de dependències sintàctiques.

Quant al processament semàntic, la semàntica lèxica s'obté d'EuroWordNet espanyol, i els rols semàntics no es consideren. En el moment que es va desenvolupar el TIPSemES, no es disposava d'eines SRL per a l'espanyol. Els treballs publicats sobre l'SRL en espanyol encara no s'han associat a programari disponible per a l'ús general (Morante, 2008).

Una anàlisi dels resultats de la versió basada en Ancora, que inclou els rols semàntics, avaluat amb les dades espanyoles del TempEval-2, es mostra en l'apartat d'avaluació en castellà de la secció 4.2.

TIPSem italià: TIPSemIT

L'italià també es va incloure en la iniciativa TempEval-2, però va obtenir una participació nul·la. No obstant això, atès que les dades en italià es van posar a disposició pública²¹, l'extensió de TIPSem era possible. Per tal de dur a terme una anàlisi fiable d'aquesta llengua, hem col·laborat amb l'institut ILC-CNR de Pisa (Itàlia).

¹⁹Una demostració en línia de la versió en text lliure de TIPSemES està disponible en: <http://gplsi.dlsi.ua.es/demos/TIMEE/>.

²⁰<http://nlp.lsi.upc.edu/freeling/>

²¹Disponible en <http://timeml.org/>

Prenent com a base l'arquitectura TIPSem, hem ampliat el nostre enfocament a l'italià (Caselli, Llorens, Saquete, & Navarro, 2011). L'extensió es va centrar en el reconeixement i la classificació d'esdeveniments, perquè no hi havia altres enfocaments per a l'italià basats en aprenentatge automàtic per a dur a terme aquestes tasques, tot i que hi ha enfocaments basats en regles per al processament de *timeX* i esdeveniments (Caselli et al., 2009, 2008). Atès que aquestes tasques no utilitzen TIMEK, l'extensió del TIMEK no és necessària. D'altra banda, a causa de la manca de recursos, només s'utilitza la semàntica lèxica en aquesta extensió.

Una de les parts més difícils d'aquest treball és el quantitat reduïda de les dades. Les dades de TempEval-2 per a l'italià no són abundants, amb 27.152 fitxes per a l'entrenament i 4.995 per a la prova. La nostra proposta manté l'estratègia de TIPSem (aprenentatge automàtic) i les característiques morfològiques generals. No obstant això, ja que no hi havia eines SRL disponibles en italià, es va utilitzar només semàntica lèxica.

Els trets semàntics de TIPSemIT s'obtenen a partir d'un lèxic d'esdeveniments derivats del lèxic SIMPLE/CLIPS (Ruimy et al., 2003). Aquest diccionari ha sigut obtingut de manera semiautomàtica a partir d'una correlació entre les classes d'esdeveniments i les entrades TimeML SIMPLE/CLIPS relacionats amb els tipus d'esdeveniments semàntics. El recurs proporciona a cada esdeveniment una o més classes TimeML. La figura 3.13 mostra una petita part del lèxic de les entrades verbals.

| | |
|------------|-----------------------------|
| dimorare | STATE |
| dimostrare | I_ACTION- OCCURRENCE- STATE |
| dipanare | I_ACTION- OCCURRENCE |
| dipartirsi | I_ACTION- OCCURRENCE |
| dipendere | OCCURRENCE |
| dipingere | I_ACTION- OCCURRENCE |
| diplomarsi | I_ACTION- OCCURRENCE |
| diradare | I_ACTION- OCCURRENCE |
| diramare | I_ACTION- OCCURRENCE |
| dire | REPORTING- OCCURRENCE |

Figura 3.13: Verbs del lèxic SIMPLE/CLIPS

El lèxic d'esdeveniments derivats hauria d'ajudar el sistema de reconeixement d'esdeveniments, ja que proporciona una llista de lemes relacionats amb esdeveniments potencials. A més, cal ajudar en la classificació dels fenòmens ambigus. Per exemple, com il·lustren els exemples 77 i 78, el verb *pensare* (CA: 'pensar') pot representar un esdeveniment de les classes TimeML *I.STATE* i *OCCURRENCE*. L'assignació de la classe correcta és estrictament depenent de l'ocurrència en el text/discurs. No obstant això,

el lèxic derivat limita l'elecció d'aquestes dues classes, cosa que hauria de millorar l'enfocament.

(77) *Marco pensa.* [Marco thinks.]
Marco <EVENT class="OCCURRENCE">pensa</EVENT>.

(78) *Marco pensa di andare a casa.* [Marco thinks to go home]
Marco <EVENT class="I_STATE">pensa</EVENT> di
<EVENT class="OCCURRENCE">andare</EVENT> a casa.

La informació semàntica té un paper primordial en l'assignació de les classes TimeML. No obstant això, no sempre és una condició necessària i suficient per a classificar-les. Altres nivells d'informació lingüística, com l'estructura argumental, poden influir en l'assignació de la classe.

L'avaluació de TIPSemIT i la seua comparació amb un enfocament basat en regles per a l'italià – Conversor Tule shortciteRobaldo-Caselli2011-TULE – es mostren en la subsecció italiana de la secció 4.2.

TIPSem xinès: TIRSemZH

Per al xinès, les tasques relacionades amb el processament d'informació temporal es troben en una fase exploratòria on calen enfocaments d'alt rendiment. De fet, no hi ha aproximacions per als esdeveniments i només unes poques per al reconeixement de *timex*. L'extensió de TIPSem és senzilla perquè, encara que no hi va haver participants en xinès, en TempEval-2, les dades TimeML van ser creades i estan disponibles públicament²² (Xue & Zhou, 2010). Per tal de dur a terme una anàlisi fiable d'aquesta llengua, hem col·laborat amb la Universitat de Xongqing (Xina).

TIPSem s'ha estès al xinès per a les tasques de reconeixement de *timex* i esdeveniments, i utilitzant només els rols semàntics com a trets semàntics. Com que el processament de la informació temporal en xinès es troba encara en una etapa primerenca i cal una anàlisi més detallada, es va abordar només l'impacte dels rols semàntics en tasques de reconeixement. TIMEK no s'utilitza en aquestes tasques i la seua extensió no va ser necessària. Hem desenvolupat un enfocament basat en aprenentatge automàtic **TIR-SemZH** (Temporal Information Recognition in Chinese), que inclou els rols semàntics i les característiques basades en la morfosintaxi.

²²<http://timeml.org/>

Pel que fa a l'expressió lingüística del temps, una de les peculiaritats del xinès mandarí és que no té morfemes de temps verbal. El temps verbal l'expressen els marcadors aspectuals i és gramaticalitzat en expressions temporals com adverbis temporals (per exemple, *demà*) o referències explícites del calendari (per exemple, *1991*).

A més de l'anglès, TimeML i la seua recent proposta d'ISO-TimeML (2007) cobreixen altres idiomes, incloent-hi el xinès ²³. L'exemple (79) mostra una frase amb l'anotació TimeML d'expressions temporals TIMEX3 i esdeveniments EVENT.

- (79) 约翰<TIMEX3>星期一</TIMEX3>
 <EVENT>来的</EVENT>。
 EN: John <EVENT>came</EVENT> on
 <TIMEX3>Monday</TIMEX3>

Tal com va ser revisat en el capítol 2, no hi ha mètodes de processament d'esdeveniments per al xinès. A més, les aproximacions de processament de *timex* han trobat problemes en el reconeixement d'expressions ambigües en el nivell morfosintàctic. Això ens motiva a adaptar TIPSem al xinès. El nostre enfocament es diferencia de les propostes actuals en diversos aspectes. Fem servir un esquema d'anotació estàndard (TimeML) i el nostre enfocament s'avalua en un conjunt de dades públic (TempEval-2). L'enfocament utilitza rols semàntics que mai s'han utilitzat en xinès per a aquestes tasques i que poden evitar les dificultats trobades en els treballs relacionats.

Igual que en altres idiomes, en els rols semàntics en xinès, capta el significat d'una frase pel que fa a com els seus arguments estan relacionats. L'exemple 80 mostra un parell de frases marcades.

- (80) 1. [他_{agent}] [一周后_{tmp}] 要参加 [比赛_{patient}]。
 EN: [He_{agent}] will attend [the competition_{patient}]
 [next week_{tmp}].
 2. [一周后_{tmp}] [他_{agent}] 要参加 [比赛_{patient}]。
 EN: [Next week_{tmp}], [he_{agent}] will attend
 [the competition_{patient}].

Els conjunts més utilitzats de rols semàntics, actualment disponibles en xinès, són Chinese-FrameNet (CFN) (You & Liu, 2005) i Chinese-PropBank (CPB) (Xue, 2008). Cada proposta se centra en un nivell diferent de granularitat.

Per a mantenir la filosofia de la semàntica TIPSem original (conjunt de rols PropBank), es va utilitzar la CPB en l'extensió al xinès.

²³http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf

Les característiques TIPSem sobre morfosintaxi i rols semàntics es van obtenir utilitzant l'eina LTP (Che, Li, & Liu, 2010). Aquesta és una plataforma de llenguatge que conté la segmentació de frases, tokenització, PoS tagging, el reconeixement d'entitats nombrades, desambiguació del sentit de la paraula, l'anàlisi de la dependència sintàctica i rols semàntics. LTP va ser alliberat al 2006 i s'ha aplicat a la investigació i en empreses internacionals en projectes de negoci real. El mòdul d'etiquetatge de rols semàntics utilitza mètodes estadístics basats en la màxima entropia, que van obtenir el primer lloc al CoNLL 2009 (Che, Li, Li, Guo, Qin, & Liu, 2009). S'obté una F1 de 77,2%, amb una eficiència del voltant dels 1.3KB/s.

TIRSemZH ha sigut avaluat i els resultats es presenten i s'analitzen en la subsecció de la secció xinesa 4.2.

3.5 Conclusions

La principal aportació de la nostra proposta als plantejaments de l'estat de la tècnica és l'aplicació de determinades característiques semàntiques al processament de la informació temporal. Hem presentat una aproximació semàntica a aquesta tasca: el sistema TIPSem. TIPSem és un enfocament híbrid (basat en aprenentatge automàtic i basat en regles), que inclou característiques semàntiques (semàntica lèxica, els rols semàntics, i la semàntica temporal), a més de les característiques morfosintàctiques.

L'arquitectura de TIPSem es divideix en tres mòduls que aborden les diferents subtasques del processament de la informació temporal: el processament de *timex* (reconeixement, classificació i normalització), processament d'esdeveniments (reconeixement i classificació), i el processament de relacions temporals (categorització).

L'enfocament es basa principalment en aprenentatge automàtic, però també inclou un recurs basat en regles (TIMEK) per a obtenir algunes característiques, així com per a normalitzar les *timex* a la norma ISO 8601. La part basada en aprenentatge automàtic inclou dues tècniques diferents; s'hi empren: camps aleatoris condicionals (CRF) i màquines de suport vectorial (SVM). Els CRF són útils per als problemes en què la classificació d'un element depèn de la classificació i les característiques dels elements adjacents: la seqüència és rellevant. Aquests s'utilitzen en el reconeixement de *timex*, el reconeixement d'esdeveniments, i la categorització de relacions. Els SVM són útils per als problemes en què la classificació d'un element és independent de la classificació dels elements adjacents: la seqüència no és rellevant. Aquests s'utilitzen en la classificació de *timex*, la classificació

d'esdeveniments, i la classificació tipus de normalització de *timex*.

La part més important del sistema és l'ús de la semàntica. Les característiques semàntiques definides per a cada tasca es resumeixen a continuació:

- **Processament de *timex***

- *Semàntica lèxica*: S'utilitzen els quatre hiperònims més generals de cada paraula com una característica del nostre model, perquè les paraules que figuren en *timex* són sovint hipònims dels conceptes en anglès ***time***, ***time_period*** o ***time_unit***, que estan entre els quatre hiperònims més generals en la jerarquia. Aquesta característica ha d'augmentar la probabilitat de representar *timex* per a paraules relacionades amb el temps, fins i tot si els seus lemes no apareixen en l'entrenament, la qual cosa hauria d'afavorir la generalització.

En la classificació de *timex* i la normalització, les característiques de semàntica lèxica s'obtenen mitjançant TIMEK, entre aquestes estan: *patró*, *granularitat d'hores*, i *indicador de repetició*. Aquestes característiques són més específiques que la semàntica lèxica general i ajuden en la diferenciació dels tipus de *timex*.

- *Rols semàntics*: S'hi empenen dues característiques – *rol (in isolation)* i *rol+lema*.

La característica rol ha d'augmentar la probabilitat de representar una *timex* per a paraules que pertanyen a rols temporals, independentment dels seus lemes (i també a disminuir les probabilitats de les paraules fora dels rols temporals). La combinació de rol+lema ha d'augmentar la probabilitat que representa una *timex* de les combinacions dels lemes i els rols temporals (p. e.: [*after the weekend* *AM-TMP*]) i també disminuir la probabilitat dels lemes no temporals que apareguen en rols temporals (p. e.: [*after the war* *AM-TMP*]). A més, aquestes característiques haurien d'ajudar en les expressions ambigües, com es mostra en (81).

- | | | |
|------|---|---------------|
| (81) | [April _{A0}] likes apples. | (female name) |
| | I went to Canada [in April <i>AM-TMP</i>]. | (timex) |
| | He saw you [last week <i>AM-TMP</i>]. | (timex) |
| | [Last week _{A0}] is my favorite song. | (song title) |

- **Processament d'esdeveniments**

- *Semàntica lèxica*: Igual que en el processament de *timex*, els quatre hiperònims més generals de cada paraula s'utilitzen com una característica del nostre model, ja que molts lemes que representen esdeveniments són hipònims dels conceptes en anglès *event*, *communication*, *perception* o *state*, i aquests es troben entre els quatre hiperònims més generals en la jerarquia. Aquesta característica ha de contribuir tant al reconeixement com a la classificació dels esdeveniments.
- *Rols semàntics*: La relació entre els rols semàntics i els esdeveniments és doble: (i) molts esdeveniments són verbs principals que tenen arguments amb diferents rols, i (ii) també molts esdeveniments poden aparèixer en diferents rols de verbs específics (p. e.: A0+happen, A1+carry_out, A2+be, etc.) i fora d'altres rols (p. e.: AM-MOD, A4, etc.). Per això, considerem les següents característiques basades en rols per a reconèixer i classificar esdeveniments: *rol* (*aïllat*), *rol+paraula*, *rol+verb*, i *rol+top4hypers*. La frase (57a) mostra com el rol A2 del verb *to be* sovint representa esdeveniments de tipus estat. Les frases (82b) i (82c) mostren per què no només el rol sinó la seua combinació amb la paraula és rellevant per a reconèixer esdeveniments. Les frases (82d) i (82e) mostren per què no només el rol sinó també la seua combinació amb el verb és rellevant. Finalment, l'esdeveniment *killing* de la frase (82f) pot ser generalitzat a altres esdeveniments similars (p. e.: *murder*, *execution*, *homicide*, etc.) usant la característica *rol+top4hypers*.

(82) a. [The tax rates _{A1}] were [low _{A2}].
b. [He _{A0}] completed [the transaction _{A1}].
c. [He _{A0}] completed [the puzzle _{A1}].
d. [He _{A0}] carried out [an experiment _{A1}].
e. [He _{A0}] ate [an apple _{A1}].
f. [The killing _{A0}] happened [this morning _{AM-TMP}].

• Processament de relacions temporals

- *Semàntica lèxica*: Per a la categorització de la relació temporal entre dos esdeveniments en una relació de subordinació, fem servir com a característica la classe TimeML d'ambdós esdeveniments. Aquesta característica, en combinació amb el temps i l'aspecte verbal, ajuda a determinar el tipus de relació temporal. En l'exemple (83), sabent que *deciding* és un LACTION i

relocating un OCCURRENCE, ajuda a determinar que *deciding* precedeix a *relocating* en el temps.

(83) We will <event eid="1"class="ACTION">decide</event> where
to <event eid="2"class="OCCURRENCE">relocate</event> the
University.

<TLINK lid=1.11"reltype="BEFORE1.leid="11.leid="2"/>

- *Rols semàntics*: Si un element de la relació (esdeveniment o *timex*) figura en un rol temporal en una subordinació, la conjunció governant de vegades és important per a determinar el tipus de relació. Conjuncions com *when*, *after*, o *while* denoten relacions temporals i aporten informació sobre la seua categoria. Per exemple, si són *when* o *while*, la categoria és més probable que siga *overlap*, mentre que si és *after* o *once*, la categoria és més probable que siga *after* – vegeu l'exemple (84). No obstant això, aquesta informació aïllada no és suficient per a categoritzar totes les relacions. Els rols semàntics són importants perquè aquestes conjuncions, quan apareixen en altres rols semàntics, poden no estar relacionades amb el temps (84c).

- (84) a. [After the Iraqi annexation of Kuwait *AM-TMP*],
oil prices rose.
b. He was rescued, [while the thieves were stealing
the safe *AM-TMP*].
c. That shop is [next to the bank we passed *AM-LOC*].

- *Semàntica temporal*: De vegades una oració conté diverses expressions temporals i esdeveniments. Si l'esdeveniment de la relació està sintàcticament vinculat a una *timex* (no considerat en la relació), la funció de la posició temporal (*valdiff*) representa si aquesta *timex* és *anterior*, *equivalent* o *posterior* en el temps a la *timex* de la relació (la diferència de valor). Com es mostra en (85), si es demana al sistema de classificar la relació de e1-t2, val la pena tenir en compte que e1 està sintàcticament relacionat amb t1 i que t1 (1999) és posterior que t2 (1998), de manera que el *valdiff* és *posterior*. No obstant això, si es demana al sistema de categoritzar e1-t1, que estan sintàcticament relacionats, aleshores, la categorització no depèn d'altres *timex* (*valdiff* = *equivalent*, per defecte).

(85) John was born_{e1} in 1999_{t1} and Sylvia was born_{e2} in 1998_{t2}.

```
valdiff (e1,t1): equal
valdiff (e1,t2): after
valdiff (e2,t1): before
valdiff (e2,t2): equal
```

Hem ampliat l'enfocament descrit anteriorment (TIPSem) de l'anglès a altres idiomes, a saber: espanyol, italià i xinès. En aquests idiomes, de manera anàloga s'apliquen trets semàntics. Un avantatge d'aquest enfocament és que, ja que està basat principalment en aprenentatge automàtic, només el preprocessament i el recurs TIMEK han de ser modificats.

El capítol següent presenta l'avaluació de TIPSem. A l'avaluació es mostra si l'ús proposat de la semàntica és útil i es mesura la seua contribució a la tasca en anglès i també des d'una perspectiva multilingüe.

Capítol 4

Avaluació

Els principals objectius d'aquesta avaluació són mesurar el rendiment del nostre enfocament semàntic, TIPSem, i mesurar i analitzar l'efecte de l'aplicació dels trets semàntics proposats, a més de la morfosintaxi, al processament de la informació temporal. Per això, TIPSem es compara amb un *baseline* no semàntic, TIPSem-B, que només utilitza les característiques morfosintàctiques.

Aquest capítol està dividit en dues seccions: (i) una avaluació principal sobre text en anglès, i (ii) una avaluació multilingüe, que inclou espanyol, italià i xinès.

La primera secció se centra en l'avaluació de la nostra proposta sobre text anglès. Aquesta és la d'avaluació principal, perquè l'anglès és actualment l'idioma per al qual més recursos de PLN estan disponibles, i al qual més esforços de recerca s'han dedicat. Aquesta avaluació se centra en la prova proposada en l'exercici d'avaluació TempEval-2 (Verhagen et al., 2010). A més, inclou una validació creuada (*10-fold cross validation*) de les dades per a analitzar la rellevància estadística de la contribució de trets semàntics.

La segona secció consisteix en una avaluació multilingüe. L'objectiu d'aquesta avaluació és comprovar si l'efecte de l'aplicació de la semàntica és comparable en tots els idiomes, així com mesurar el rendiment de TIPSem en cada idioma.

4.1 Avaluació sobre text en anglès

La nostra aproximació (TIPSem) i el *baseline* (TIPSem-B) s'avaluen per a mesurar-ne el rendiment, i es comparen per a analitzar l'efecte de la semàntica. D'altra banda, per a estudiar la contribució de cada tipus de

semàntica de manera individual, i per mesurar quin tipus de semàntica és més beneficiós per a cada subtasca, s'inclouen tres enfocaments intermedis que usen només part del coneixement : TIPSem-SR (rols semàntics), TIPSem-LS (semàntica lèxica), i TIPSem-TS (semàntica temporal). La figura 4.1 descriu com s'ha mesurat la influència de la semàntica.

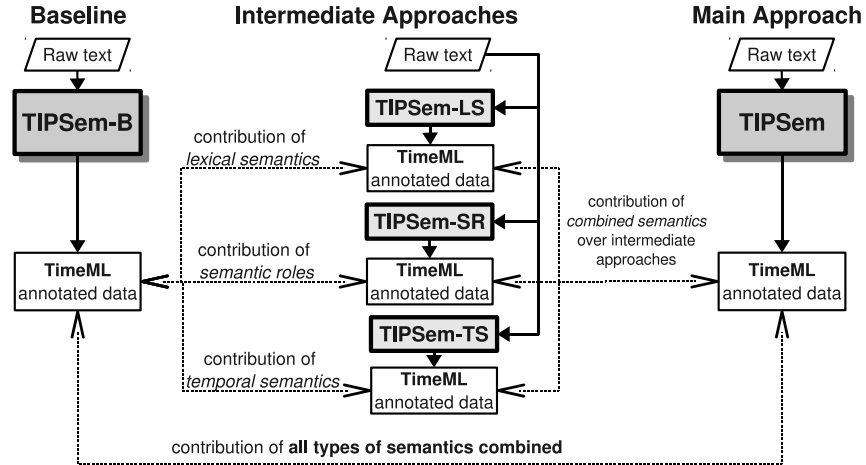


Figura 4.1: Mesurant la contribució de la semàntica

En les següents subseccions descrivim l'entorn de l'avaluació, els resultats i la seva anàlisi, i la comparació amb l'estat de la qüestió.

Entorn d'avaluació

El marc d'avaluació consisteix en un conjunt de dades i un criteri d'avaluació.

Conjunt de dades

Els enfocaments presentats s'han entrenat i provat usant el conjunt de dades públic del TempEval-2¹. Aquestes dades són el nostre test principal, ja que permeten la comparació dels nostres resultats amb els obtinguts per altres sistemes en aquest recent exercici d'avaluació –vegeu l'apèndix D.

La taula 4.1 resumeix les estadístiques de les dades del TempEval-2. Aquesta informa sobre la densitat dels elements TimeML en el text, així com els tipus i classes més populars. Es pot observar que hi ha molts més esdeveniments (6186) que *timex* (1135), la classe d'esdeveniment més freqüent

¹<http://timeml.org/site/timebank/timebank.html>

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

és l'*occurrence* (3511), i el tipus més freqüent de *timex* és *date* (939). Pel que fa a les relacions temporals, el tipus més popular de les relacions entre els esdeveniments són les de *main-events* i *subord-events* (1861 de 5439). Finalment, mirant cada tipus de relació, en *event-timex* la categoria més popular és *overlap* (557 de 1024), per a l'esdeveniment-DCT és *before* (467 de 830), per als *main-events overlap* (731 de 1724), i *subord-events before* (651 de 1861).

| Set | Docs | Words | Element: classes |
|-------------------------|------|-------|---|
| Train | 162 | 53K | TIMEX(1052) : date(870), duration(153), time(15), set(12) EVENT(5688) : occurrence(3220), reporting (838), i_action(530), i_state(440), state(426), aspectual(196), perception(38) TLINK event-timex(959) : overlap(516), before(172), after(221), before-or-overlap(15), overlap-or-after(16), vague (19) TLINK event-dct(640) : overlap(188), before(335), after(63), before-or-overlap(23), overlap-or-after(12), vague (19) TLINK main-events(1587) : overlap(652), before(403), after(279), before-or-overlap(61), overlap-or-after(51), vague (141) TLINK subordinated-events(1721) : overlap(518), before(600), after(299), before-or-overlap(111), overlap-or-after(89), vague (104) |
| Test (entities) | 9 | 5K | TIMEX(81) : date(69), duration(8), time(4), set(0) EVENT(498) : occurrence(291), reporting (66), i_action(53), i_state(27), state(26), aspectual(31), perception(4) |
| Test (relations) | 11 | 5K | TLINK event-timex(65) : overlap(41), before(8), after(10), before-or-overlap(1), overlap-or-after(2), vague (3) TLINK event-dct(190) : overlap(45), before(112), after(24), before-or-overlap(4), overlap-or-after(2), vague (3) TLINK main-events(137) : overlap(79), before(26), after(13), before-or-overlap(5), overlap-or-after(4), vague (10) TLINK subordinated-events (140) : overlap(47), before(51), after(21), before-or-overlap(17), overlap-or-after(3), vague (1) |

Taula 4.1: Estadística de les dades TimeML en anglès (TempEval-2)

La taula 4.2 informa sobre el grau d'acord entre els anotadors humans (*inter-annotator agreement*) en el reconeixement dels elements TimeML, i

la taula 4.3 mostra la precisió i cobertura mitjanes (P&R) i el valor kappa en els atributs TimeML. L'acord entre els anotadors, la P&R, i el valor kappa només van ser calculats sobre l'anotació del TimeBank², però com que TimeBank i TempEval-2 són en la seua majoria equivalents, aquests valors serveixen com a avaluació indirecta.

Aquesta informació suggereix indirectament la complexitat de les diferents tasques, en termes d'acord amb l'anotació humana. D'aquestes dades, es pot observar que, per anotadors humans, els esdeveniments i relacions temporals són més complexes de processar que les *timex*.

| TimeML tag | agreement |
|--------------|-----------|
| TIMEX | 0,83 |
| EVENT | 0,78 |

Taula 4.2: Acord entre anotadors en TimeML anglès

| TimeML tag | P&R | Kappa |
|----------------------|------|-------|
| TIMEX.type | 1.00 | 1.00 |
| TIMEX.value | 0,90 | 0,89 |
| EVENT.class | 0,77 | 0,67 |
| TLINK.relType | 0,77 | 0,81 |

Taula 4.3: P&R i Kappa del TimeML anglès

El conjunt de dades TempEval-2 en anglès es basa en el corpus TimeBank (Pustejovsky et al., 2003) i suposa una revisió de la seua anotació. Comprèn exactament els mateixos documents, excepte *wsj_0685*. Els documents provenen d'una varietat d'articles de notícies del programa Extracció Automàtica de Contingut (ACE) i de PropBank (TreeBank2). Els articles del primer són transcripcions de notícies i els subministrats per PropBank són del *Wall Street Journal*. Tots dos són part de les dades de *Penn Treebank*.

Criteri i mesures d'avaluació

El criteri utilitzat en TempEval-2 (SemEval-2010) s'ha aplicat en aquesta tesi per a afavorir la comparació.

²<http://timeml.org/site/timebank/documentation-1.2.html>

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

D’una banda, les tasques de reconeixement de *timex* i esdeveniments s’han avaluat mitjançant les mesures següents:

$$\text{precision (presició)} = \frac{\text{true_positives}}{\text{true_positives} + \text{false_positives}}$$

$$\text{recall (cobertura)} = \frac{\text{true_positives}}{\text{true_positives} + \text{false_negatives}}$$

$$\mathbf{F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

on *true_positives* són el nombre de paraules que formen part de l’extensió d’una entitat temporal en el fitxer clau (anotat per humans) i en l’eixida del sistema, *false_positives* són el nombre de paraules que són part de l’extensió de les entitats temporals anotades pel sistema però que no són al fitxer clau, i *false_negatives* són el nombre de paraules que formen part de l’extensió de les entitats temporals en el fitxer clau, però no en l’eixida del sistema.

D’altra banda, la resta de tasques, és a dir, la classificació de *timex* i esdeveniments, la normalització i la categorització de relacions temporals s’avaluen utilitzant la mesura:

$$\text{accuracy (exactitud)} = \frac{\text{correctly_classified_instances}}{\text{correctly_recognized_instances}}$$

Durant el desenvolupament d’aquesta tesi, hem observat el següent problema en el criteri d’avaluació TempEval-2: l’exactitud es va calcular sobre els casos reconeguts correctament per cada sistema en particular, de manera que hem canviat el nom a **partial.accuracy** (PA). La puntuació depèn de la quantitat de casos reconeguts correctament. Per tant, dos sistemes que reconeixen un conjunt diferent de *timex* i esdeveniments no es poden comparar. L’exemple (86) mostra el problema.

(86) The gold standard corpus contains: **81** *timex*

System A recognizes 1 *timex*
and **classifies 1 correctly** \rightarrow **partial.accuracy** = $\frac{1}{1} = 1.0$

System B recognizes 40 *timex*
and **classifies 32 correctly** \rightarrow **partial.accuracy** = $\frac{32}{40} = 0,8$

Problems: (i) each system classified a different set of *timex*, and (ii) the second system classifies more *timex* correctly and obtains a lower score. The **scores** of systems A and B are **not comparable**.

Aquest problema afecta la classificació de *timex* i esdeveniments, a la normalització, i a la categorització. No obstant això, en la categorització,

com que tots els sistemes ³ utilitzen els esdeveniments i *timex* clau, poden ser comparats entre si en aquesta tasca.

En la nostra avaluació, a més dels resultats de *partial_accuracy* (compatible amb TempEval-2), presentem els nostres resultats de classificació i normalització calculats sobre totes les *timex* i esdeveniments del corpus clau, no només sobre els reconeguts per un sistema particular. Això proporciona una puntuació d'exactitud comparable. Hem anomenat a la nova mesura **complete_accuracy** (CA). L'exemple (87) mostra com d'aquesta manera el problema anterior s'evita.

(87) The gold standard corpus contains: **81** *timex*

System A recognizes 1 *timex* (but is asked to classify the **81** *timex*)
and classifies **40** correctly \rightarrow **complete_accuracy** = $\frac{40}{81} = 0,49$

System B recognizes 40 *timex* (but is asked to classify the **81** *timex*)
and classifies **60** correctly \rightarrow **complete_accuracy** = $\frac{60}{81} = 0,75$

Now **the accuracy can be compared** because both systems are evaluated over the same *timex* (the 81 gold standard *timex*).

Resultats i anàlisi

Els resultats de cada entitat temporal es presenten per separat (*timex*, esdeveniments i relacions temporals). A més, aquests han sigut analitzats tenint en compte les subtasques relacionades amb cada entitat (per exemple, el reconeixement, la classificació, etc.). Per a simplificar la lectura, les millors puntuacions F1 i d'exactitud es destaquen en negreta en les taules de resultats.

Processament de *timex*

Quant a processament d'expressió temporal, es diferencien tres tasques: reconeixement, classificació i normalització.

Reconeixement de *timex*

El reconeixement de *timex* es defineix com la delimitació exacta d'aquestes expressions en el text. La taula 4.4 mostra els resultats obtinguts en aquesta tasca en anglès sobre les dades de la prova TempEval-2. Els resultats s'analitzen exemplificant (i) la contribució de la semàntica lèxica, (ii) la contribució dels rols semàntics, i (iii) els errors observats.

³Excepts TRIOS/TRIPS.

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

| Approach | precision | recall | F1 |
|---------------|-----------|--------|-------------|
| TIPSem-B | 0,89 | 0,68 | 0,77 |
| TIPSem-LS | 0,91 | 0,79 | 0,85 |
| TIPSem-RS | 0,91 | 0,76 | 0,83 |
| TIPSem | 0,93 | 0,84 | 0,88 |

Taula 4.4: Reconeixement de *timex* (anglès)

El rendiment de TIPSem en aquesta tasca va arribar a un F1 alt (0,88). En anglès, l'aplicació de la semàntica lèxica, els rols semàntics, i la seua combinació van obtenir una reducció de l'error en F1 sobre el *baseline* del 35%, 26%, i 59% respectivament. La combinació d'ambdós tipus de coneixement semàntic obté la major reducció d'error (59%), la qual cosa indica que la contribució dels diferents tipus de semàntica és complementària.

La rellevància de la millora en F1 s'ha estudiat aplicant un t-test sobre una divisió en 10 mostres del corpus. La reducció mitjana de l'error en F1 obtinguda per TIPSem sobre TIPSem-B (26%) és estadísticament significativa, amb un nivell de confiança del 99,5%⁴. Es poden trobar més detalls a l'apèndix C.

A continuació s'analitza i es discuteix la contribució de cada tipus de semàntica i els errors generals que s'han trobat:

- **Contribució de la semàntica lèxica:** la presència de classes semàntiques relacionades amb el temps en algunes paraules contingudes en les *timex* contribueix a que siguen creconegudes correctament. Per exemple, *December* obté la classe semàntica *time_period*, i l'hora *hour* obté *time_unit*. Amb aquesta informació, l'enfocament s'assabenta que l'obtenció d'aquests valors en la característica *top4hypers* augmenta la probabilitat de trobar una *timex*. L'exemple (88) mostra una *timex* que el *baseline* ignora, però que TIPSem-LS reconeix correctament.

(88) TempEval-2 test set - file: V0A19980501.1800.0355 - sentence: 1
He and Palestinian leader Yasser Arafat meet separately Monday with
US Secretary of State.

El *baseline* ignora la *timex Monday* perquè el seu context morfo-sintàctic no és comú en les dades d'entrenament, i per tant, la probabilitat del model en context no és prou alt per a reconèixer la *timex*. No

⁴t(9)=4.18 (one-tail $p=0,005$). p : probabilitat d'obtenir la millora per casualitat.

obstant això, utilitzant la semàntica lèxica, *Monday* s'obté el valor *time-period => quantity => measure => abstraction* per a la característica *top4hypers*, el qual és molt comú entre les *timex* de l'entrenament, i incrementa la probabilitat fins al punt que reconeix *Monday* com a *timex*.

Tot i que l'aplicació de la semàntica lèxica ha sigut positiva per a la tasca, en alguns casos ha disminuït el rendiment (vegeu (89)).

(89) TempEval-2 test set - file: WSJ900813-0157 - sentence: 2
James Baker, speaking in ABC News *This Week* said the Kuwaiti [...]

En (89), *This Week* no va ser reconeguda com a *timex* pel *baseline* perquè *week* seguida d'una cometa doble no es troba com a *timex* en l'entrenament. No obstant això, si s'aplica la semàntica lèxica, *week* és *time-period => quantity => measure => abstraction*, la qual cosa incrementa la probabilitat de ser una *timex* com s'ha esmentat anteriorment. Això és normalment útil, però en aquest cas *This Week* és el nom de la secció de les notícies d'ABC i no una *timex*. Per tant, la semàntica lèxica és insuficient per a resoldre aquest tipus d'ambigüitat.

- **Contribució dels rols semàntics:** Aquesta se centra principalment en el rol temporal i els arguments relacionats amb el temps dels verbs específics (p. e.: el rol A2 de *to last*). El sistema aprèn que les paraules que figuren en els rols temporals són més propenses que altres a ser *timex*. Anàlogament, les paraules fora dels rols temporals tenen menys probabilitats de ser *timex*. *This Week* com una *timex*. Com que aquesta expressió té un rol A1 que normalment no conté *timex* en l'entrenament, la probabilitat de reconeguen la *timex* disminueix. Per tant, l'error de TIPSem-LS es resol per TIPSem mitjançant l'ús dels rols.

Els rols semàntics també ajuden en (90). El nombre *1360*, l'ha marcat com una *timex* el *baseline* però el problema es resol mitjançant l'ús dels rols semàntics.

(90) TempEval-2 test set - file: wsj.0505 - sentence: 6
The number of stores would increase to 1450 stores from 1360.

Les aproximacions no basades en rols fallen perquè la seqüència de *from number* té una alta probabilitat de representar una *timex*. No obstant això, els rols semàntics assenyalen que per al verb *increase*, *the number of stores*, és la cosa que s'incrementa (rol A1), *to 1450 stores* és el punt

final d'aquest augment (rol A4), i *from 1360*, el punt de partida (rol A3), però res relacionat amb el temps.

Un altre dels avantatges dels rols semàntics és que delimiten els arguments temporals d'un predicat. Això és útil per a detectar les expressions temporals multiparaula. La *timex* en (91) va ser limitat incorrectament pel *baseline* i per TIPSem-LS – només la part *the week* va ser reconeguda. No obstant això, l'expressió va ser limitada correctament pels enfocaments que inclouen els rols perquè *the end of this week* està contingut al rol AM-TMP⁵.

- (91) TempEval-2 test set - file: wsj_0073 - sentence: 0
 [...] and the company will begin mailing materials to shareholders
 at the end of the week.

En alguns casos, com en (92), l'aplicació dels rols semàntics redueix el rendiment.

- (92) TempEval-2 test set - file: APW19980306.1001 - sentence: 13
 [...] these sites could only be visited by diplomats as laid down by the
Feb. 23 accord signed by [...]

En (92), *Feb. 23 (accord)* no té el rol temporal, ja que és l'agent del predicat encapçalat per *laid down*. Per tant, TIPSem falla en reconèixer-lo com a *timex*. Les *timex* representades per modificadors de substantius són difícils de reconèixer. Alguns casos, com *I live on November 6th street on November 6th* no té un rol temporal, són correctament descartats com a *timex* –tenir en compte el nom del carrer com a *timex* portaria a una interpretació temporal errònia de la frase. No obstant això, el reconeixement d'una *timex* en forma de modificador de substantiu no depèn del rol, sinó del substantiu al qual modifica. Per exemple, els noms d'acords és probable que siguin *timex* reals, però els carrers no.

- **Errors trobats.** Finalment, deixant de banda la contribució de la semàntica, s'han analitzat els errors que impedeixen a TIPSem obtenir un rendiment del 100%. Aquests errors es poden resumir en tres tipus:
 - *timex modificadors de substantius*: TIPSem ha fallat algunes expressions temporals que apareixen com a modificadors de noms (p. e.: 10-hour tour, third quarter loan) per la seua complexitat

⁵També *at* està dins l'AM-TMP, però el model aprèn que les preposicions governants dels arguments s'exclouen de les *timex*.

- no són molt comuns en les dades d’entrenament i no apareixen com a part de rols temporals, com s’ha explicat anteriorment.
- *Errors i inconsistències de l’anotació clau*: TIPSem ha anotat algunes *timex* que no van ser anotades pels anotadors humans en el conjunt de dades TempEval-2 (p. e.: *the first days, a few days*). Des del nostre punt de vista, aquestes són majoritàriament *timex* reals que no han sigut anotades pels anotadors humans perquè els han passat per alt (errors humans). A més, trobem alguns casos inconsistents –expressions que apareixen anotades com a *timex* de vegades, però altres vegades no (p. e.: *previously, now, currently*). Cal no oblidar que l’acord entre anotadors humans per al reconeixement de *timex* és del 83%.
- *Errors d’etiquetatge de rols semàntics (SRL) errors*: Alguns errors han sigut causats per rols temporals incorrectament etiquetats com a arguments numerats (p. e.: [next week _{A1}]). En particular, el rendiment de l’eina d’SRL utilitzada en l’etiquetatge de rols temporals és del 77% d’F1.

En poques paraules, la contribució de la semàntica en el reconeixement de *timex* es podria resumir de la manera següent. La semàntica lèxica i els rols semàntics introdueixen avantatges complementaris en l’enfocament, incrementant la seua capacitat de generalització. L’aplicació de la semàntica lèxica, augmenta la probabilitat de representar una *timex* per als conceptes que es relacionen semànticament amb el temps en WordNet (encara que no apareguen a l’entrenament). Mitjançant l’ús de rols semàntics, l’enfocament augmenta la probabilitat de reconèixer una *timex* per als arguments que tenen un rol temporal o un argument relacionat amb el temps d’un verb (per exemple, el rol A2 de *to last*), la qual cosa és útil els casos ambigus com *ABC News This Week, 1360*, etc. Finalment, en general, TIPSem és un sistema competent en aquesta tasca (0,88 F1).

Classificació de *timex*

Aquesta tasca es defineix com l’obtenció del valor de l’atribut *type* d’una *timex* reconeguda prèviament, és a dir, classificar una *timex* com a pertanyent a un dels tipus de *timex* segons TimeML (*i. e.*, *date, time, duration, set*). En TempEval-2, aquesta tasca es va mesurar pel que es refereix a *partial accuracy* (la quantitat de casos correctament classificats dels casos reconeguts correctament per cada sistema). Com s’ha discutit en la secció 4.1, per tal de comparar els diferents enfocaments, s’introdueix la mesura

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

complete_accuracy que es calcula sobre totes les *timex* corpus. La taula 4.5 mostra els resultats obtinguts en aquesta tasca en anglès en les dades del TempEval-2.

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,94 | 0,93 |
| TIPSem | 0,98 | 0,98 |

Taula 4.5: Classificació de *timex* (anglès)

El rendiment més alt obtingut en aquesta tasca és de 0,98 en complete_accuracy. Centrant-se en complete_accuracy, les puntuacions oscil·len del 0,93 de TIPSem-B al 0,98 de TIPSem. Com que els rols semàntics no estan inclosos en la nostra aproximació de classificació de *timex*⁶, els resultats obtinguts indiquen que l'ús de la semàntica lèxica (xarxes semàntiques i TIMEK) és útil per a aquesta tasca (+0,05).

s'ha demostrat que la reducció mitjana d'error relativa al *baseline* obtinguda per la semàntica lèxica en una avaluació creuada (11%) és significativa amb una confiança del 97,5% (vegeu l'apèndix C).

- **Contribució de la semàntica lèxica:** La semàntica lèxica ha ajudat a classificar correctament el tipus *time*. En particular, la contribució se centra en les característiques de TIMEK (p. e.: *pattern*, *time_granularity*). Això és així perquè, per exemple, les dates (p. e.: *this year*), la durada (p. e.: *15 years*), i els temps (p. e.: *this morning*) contenen lemes (p. e.: *year*, *morning*), que comparteixen els quatre primers hiperònims. Només quan s'apliquen les característiques de TIMEK als exemples anteriors, apareixen algunes diferències en el seu patró (*this_TUnit*, *NUM_TUnit*, *this_TOD*) i granularitat (*false*, *false*, *true*).

La majoria de les *timex* són dates; per tant, el *baseline* mostra una alta probabilitat de classificar les *timex* com a data. TIPSem-LS millora el *baseline* especialment en la classificació del tipus *time* (hora) (p. e.: *night* i *morning*).

- **Errors trobats:** Només dos elements van ser classificats incorrectament per la millor aproximació (TIPSem-LS). L'exemple (93) mostra aquests dos errors.

⁶Vam comprovar que, si els rols semàntics s'inclouen, el rendiment baixa en -0,02. Això és perquè la majoria de les *timex* apareixen en rols temporals i són *dates*. I per tant, expressions com *last night* (type=*time*) es classifiquen com a *date*.

- (93) TempEval-2 test set - file: wsj_0505 - sentence: 5
 a. In the fiscal year [...], the company reported sales of [...]

- TempEval-2 test set - file: wsj_0586 - sentence: 1
 b. London sales were initially depressed by overnight losses [...]

En la frase (93a), la *timex duration* subratllada va ser classificada erròniament com una *date*, i en l'oració (93b), la *timex time* subratllada va ser classificada erròniament com una *duration*. Des del nostre punt de vista, aquestes *timex* són tant duratives com de data/hora. *the fiscal year* es refereix a una durada d'un any i a l'interval de dates específic 1988-1989; per tant, en TimeML, ha de ser anotada com a data com TIPSem ha fet. *overnight* es refereix a un període de nit difús, però assignat a una nit específica. No obstant això, l'actual definició de tipus TimeML només permet donar un tipus a una *timex*⁷.

En poques paraules, l'anàlisi dels resultats en la classificació de *timex* es pot resumir en dues idees: (i) els rols semàntics no són útils per a la classificació de *timex*, i (ii) la semàntica lèxica obtinguda de TIMEK produeix una reducció significativa d'error. Donats aquests resultats, la millor solució a aquest problema és utilitzar només la semàntica lèxica.

Normalització de *timex*

La normalització de *timex* consisteix a obtenir el valor ISO 8601 donada una *timex*. Tal com hem definit en la secció 3.2.1, el nostre enfocament porta a terme aquesta tasca en un procés de dos passos: (i) una fase basada en aprenentatge automàtic que determina quin tipus de funció de normalització cal aplicar-hi, i (ii) una fase basada en regles que aplica la funció de normalització corresponent.

La taula 4.6 mostra els resultats obtinguts. Com que només s'utilitza la semàntica lèxica, només es mostra TIPSem (TIPSem-LS i TIPSem són equivalents). A més, com a segon pas, es basa en regles; no hi ha *baseline* per a aquest pas.

La millora obtinguda en *complete_accuracy* es deu al primer pas, on TIPSem-B obté un resultat de 0,91 i TIPSem, de 0,98. La reducció mitjana de l'error relatiu sobre el *baseline* (19%) obtingut en una avaluació creuada és estadísticament significativa amb una confiança del 99,5% (vegeu l'apèndix C).

⁷TimeML gestiona aquests casos, incloent-hi etiquetes *timex* buides amb atributs extra (*anchor_id*, *begin_point*, etc.), però això podria ser més clar si hi incloguérem *timex* de tipus doble.

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

| Approach | step 1 | step 2 | partial_accuracy | complete_accuracy |
|---------------|--------|--------|------------------|-------------------|
| TIPSem-B | 0,91 | - | 0,81 | 0,75 |
| TIPSem | 0,98 | 0,84 | 0,84 | 0,81 |

Taula 4.6: Normalització de *timex* (anglès)

- **Contribució de la semàntica lèxica:** Aquesta se centra en el primer pas. La traducció del les paraules de l'expressió temporal en un patró (característica *pattern*) ajuda a fer que el model trobe la normalització que s'ha d'obtenir en el segon pas. Els casos erronis de TIPSem-B que van ser resolts per TIPSem són *timex* la normalització dels quals és PERIOD i en els quals el *baseline* falla en utilitzar ISO_function, com per exemple els que es mostren en (94).

(94) TempEval-2 test set - file: APW19980306.1001 - sentence: 0
 2 months
 TempEval-2 test set - file: wsj_0586 - sentence: 7
 2 weeks

En l'exemple (94), l'ús de la característica patró de TIMEK, que en aquests casos (*2 months* i *2 weeks*) obté *NUM_TUNIT*, és útil per a generalitzar que aquest tipus d'expressions es normalitzen amb l'aplicació de la funció PERIOD (p. e.: *P2M*, *P2W*).

- **Errors trobats:** Es distingeixen els errors trobats en el pas u i els errors trobats en el pas dos.

Al pas u, només hi ha dos casos erronis –vegeu l'exemple (95).

(95) TempEval-2 test set - file: wsj_0505 - sentence: 5
 a. In the fiscal year [...], the company reported sales of [...]
 TempEval-2 test set - file: APW19980306.1001 - sentence: 6
 b. [...] Ritter arrived on Thursday for an inspection likely to last over a week

Tots dos són *timex* duratives, perquè es normalitzen com a períodes (P1Y i P1W). No obstant això, TIPSem els va classificar com que s'havien de normalitzar com ISO_function. Una vegada més veiem el problema explicat en la classificació de *timex*: una *timex* amb dos tipus (*duration* i *date*).

En el segon pas, els errors són causats per ignorar els modificadors de les *timex*, pel temps futur del verb *to be*, i per les *timex* que són relatives a altres *timex* en lloc de ser-ho a la DCT. Per exemple:

En (96), TIPSem ha normalitzat les expressions sense tenir en compte el modificador *end* i, per tant, l'error consisteix a no fer servir l'últim dia del mes o de l'any en la normalització.

- (96) TempEval-2 test set - file: WSJ900813-0157 - sentence: 53
Pentagon officials say that the objective is to put 40,000 troops in the region by the end of the month.

TempEval-2 test set - file: wsj.0586 - sentence: 43
The percentage change is since *year-end* [...]

En (97), TIPSem falla en la normalització de les expressions, perquè el verb principal de les sentències està en temps passat. No obstant això, com que les *timex* són modificades per la construcció verbal *to be + past participle*, la direcció temporal és futura.

- (97) TempEval-2 test set - file: WSJ900813-0157 - sentence: 53
They said they expected the transaction to be completed by Dec. 15.

TempEval-2 test set - file: wsj.0505 - sentence: 3
He added the guidelines to be published in *November* [...]

TempEval-2 test set - file: wsj.0586 - sentence: 24
Dealers placed heavy buy orders in the morning to start the first trading day.

En (98), TIPSem ha normalitzat malament aquestes *timex* relatives, ja que ho ha fet respecte a la DCT. El problema és que, a diferència d'algunes *timex* relatives al temps de l'escriptura (DCT), aquestes són relatives a una referència de temps anterior (una altra *timex*). En aquests casos, cal conèixer sobre la referència anafòrica. Per exemple, en (98a) *a year ago* no es refereix a *last year* (1988) sinó a *last third quarter* (1988-Q3). El lector és conscient d'això perquè el context del discurs mostra la relació. En (98b), ens trobem amb un altre tipus de relació anafòrica entre *timex the previous Friday*. No es refereix a *last Friday* (1989-10-27), sinó al divendres anterior al fet que es reporta en el text (1989-10-20).

- (98) TempEval-2 test set - file: wsj_0527 - sentences: 5,9,11
a. Crossland reported a third-quarter loss of \$175.5 million, compared with a year ago.

TempEval-2 test set - file: wsj_0586 - sentence: 4
b. The Financial Times 100-share index shed 47.3 points to close at 2082.1 , down 4.5% from the previous Friday.

A part d'aquests tipus de *timex* anafòrica del TempEval-2, hi ha altres tipus complexos com referències temporals en l'estil indirecte o diàleg, com es mostra en (99).

- (99) Today is the 17th of December and he still feels tired.
Yesterday, I went to the party, he said yesterday.

En aquest exemple, *yesterday* en estil indirecte o de diàleg (subratllat) no ha de ser normalitzat de la mateixa forma que el *yesterday* en estil directe. Suposant que el temps d'escriptura és 17 de desembre, el primer *yesterday* (subratllat) es refereix a 15 de desembre, mentre que el segon es refereix al 16 de desembre. Això assenyala que un sistema de processament de la informació temporal ha de manejar les característiques del discurs com, per exemple, estar al corrent del fet que una *timex* es troba dins d'un diàleg i, si és així, saber en quin moment es va pronunciar aquest discurs. TIPSem encara no és capaç de normalitzar les *timex* anafòriques.

Finalment, en (100), *overnight* es va normalitzar incorrectament perquè l'expressió no estava inclosa en TIMEK en el moment de l'avaluació. Després d'aquesta anàlisi, *overnight* va ser inclosa en TIMEK.

- (100) TempEval-2 test set - file: wsj_0586 - sentence: 1
London sales were initially depressed by overnight losses [...]

En resum, la normalització de les expressions temporals es pot resoldre computacionalment amb un rendiment acceptable, per sobre de 0,80. No obstant això, sobrepassar aquesta barrera és complicat, ja que són necessàries la resolució de l'anàfora i la capacitat de processament del discurs.

Processament d'esdeveniments

Quant a processament d'esdeveniments, hem diferenciat dues tasques: reconeixement i classificació.

Reconeixement d'esdeveniments

El reconeixement d'esdeveniments es defineix com la delimitació exacta de les expressions d'esdeveniments en el text. La taula 4.7 mostra els resultats obtinguts en aquesta tasca en anglès sobre les dades de TempEval-2.

| Approach | precision | recall | F1 |
|---------------|-------------|-------------|-------------|
| TIPSem-B | 0,82 | 0,80 | 0,81 |
| TIPSem-LS | 0,81 | 0,86 | 0,83 |
| TIPSem-RS | 0,82 | 0,86 | 0,84 |
| TIPSem | 0,82 | 0,87 | 0,85 |

Taula 4.7: Reconeixement d'esdeveniments (anglès)

L'aplicació de la semàntica lèxica, els rols semàntics, i la seua combinació van obtenir una reducció d'error en F1 d'11%, 16% i 21% en relació amb el *baseline*, respectivament. Igual que en el cas de *timex*, l'aplicació de la semàntica lèxica i els rols semàntics per separat produeix una millora sobre el *baseline*, i la combinació d'ambdós tipus de coneixement semàntic obté la màxima reducció de l'error (21%), la qual cosa indica que la seua contribució és complementària.

La reducció mitjana de l'error introduït per la semàntica (11%) ha demostrat ser estadísticament significativa amb un nivell de confiança del 99,5% en una validació creuada.

- **Contribució de la semàntica lèxica:** Quan la característica *top4hypers* obté valors relacionats amb *event* o *state*, augmenta la probabilitat que una paraula siga un esdeveniment. Això afavoreix la capacitat de generalització d'aquest enfocament en casos com (101).

(101) TempEval-2 test set - file: APW19980306.1001 - sentence: 0
An American leader of a UN weapons inspection team resumed work
in Iraq Friday.

(102) This is your work.
There are opportunities to work abroad.

El nom *work* és un esdeveniment en (101), però no en altres situacions, com en (102). Atès que l'entrenament té més casos de *work* que no són els esdeveniments, el *baseline* mostra una alta probabilitat

de descartar-lo com a esdeveniment. Amb la semàntica lèxica, l'enfocament reconeix correctament *work* com un esdeveniment en (101), a causa del fet que el seu valor de *top4hypers* (*act* => *event* => *psychological_feature* => *abstraction*) és molt popular entre els esdeveniments nominal de la formació, especialment aquells precedits pel lema *resume*, i, per tant, la probabilitat de reconèixer *work* com a esdeveniment augmenta. Això afavoreix la generalització de l'enfocament per als esdeveniments nominals (p. e.: *inspection*, *occupation*, *negotiations*, *invasion*).

- **Contribució dels rols semàntics:** Aplicant els rols, el nombre d'esdeveniments reconegut correctament augmenta. Això és degut al fet que (i) certs rols (p. e.: A1 i A2) d'alguns verbs tenen més probabilitats de contenir esdeveniments, i (ii) els verbs principals, que normalment són esdeveniments, són identificats com a nuclis dels predicats per l'eina d'SRL. Per exemple, en (103a) i (103b), els esdeveniments *shakeup* i *inspection* no són detectats pel *baseline* ni pel TIPSem-LS, però es reconeixen correctament mitjançant l'ús de rols semàntics.

(103) a.TempEval-2 test set - file: wsj_0586 - sentence: 38

British shakeup was widely cited by the declines.

b.TempEval-2 test set - file: APW19980306.1001 - sentence: 6

They had been allowed to carry out an inspection.

El *baseline* (TIPSem-B) falla perquè aquests lemes no apareixen en l'entrenament. La semàntica lèxica (TIPSem-LS) augmenta la seua probabilitat de ser esdeveniments, però no prou perquè el model els reconega. No obstant això, quan els rols semàntics s'apliquen augmenta més encara la probabilitat. Aquests esdeveniments tenen el rol A1 i molts esdeveniments nominals apareixen en l'entrenament amb aquest rol; per tant, aquests són reconeguts per TIPSem. En (103b), *inspection* té el rol A1 per al verb *to carry out*, que normalment defineix esdeveniments (p. e.: *carry out* + *action/process*).

Un altre dels avantatges dels rols semàntics és que identifiquen els verbs principals, que normalment representen esdeveniments. Aquests verbs són els elements que governen els predicats. L'exemple (104) mostra un esdeveniment errat pel *baseline* i per l'enfocament de la semàntica lèxica, però correctament reconegut per TIPSem.

(104) TempEval-2 test set - file: V0A19980501.1800.0355 - sentence: 1

He and Palestinian leader Yasser Arafat meet separately Monday with US Secretary of State.

En aquest exemple, *meet* com a verb principal dels arguments del predicat A0 (*He and Palestinian leader Yasser Arafat*), AM-MNR (*separately*), AM-TMP (*Monday*) i A1 (*with US Secretary of State*). La majoria dels verbs principals són esdeveniments i el model TIPSem aprèn això. El *baseline* no, perquè el lema *meet* amb PoS *VBP* no es troba en l'entrenament com un esdeveniment, i TIPSem-LS falla perquè la jerarquia d'hiperònims verbals en WordNet és limitada i el model dóna més probabilitat a les característiques del *baseline*.

- **Errors trobats:** Hem dut a terme un estudi específic dels errors de TIPSem en el reconeixement d'esdeveniments (Llorens et al., 2010a). La taula 4.8 mostra els resultats de la nostra aproximació per categoria gramatical.

| | | Precision | Recall | F1 |
|-------------|-------------|-----------|--------|-------|
| Recognition | <i>Verb</i> | 91.56 | 92.15 | 91.33 |
| | <i>Noun</i> | 72.67 | 48.26 | 58.42 |
| | <i>Adj.</i> | 66.78 | 38.09 | 48.35 |

Taula 4.8: Reconeixement d'esdeveniments (detall per categoria gramatical)

Es pot observar que els millors resultats s'obtenen en el reconeixement dels esdeveniments verbals, seguit pels nominals i adjectivals. Això coincideix amb l'acord entre els anotadors humans⁸, i assenyala que els esdeveniments nominals i adjectivals són més difícils de reconèixer.

En particular, la semàntica evita els errors del *baseline* en paraules de diferents categories gramaticals, especialment reduït els errors en els esdeveniments nominals que es concentren al voltant del 70% de la contribució total. Això és important, ja que demostra que la semàntica és útil per a reconèixer els tipus d'esdeveniments complexos, com ara esdeveniments nominals, en els quals els enfocaments basats únicament en la morfosintaxi fallen.

En resum, la informació semàntica és útil per a abordar el reconeixement dels esdeveniments. La semàntica lèxica afavoreix la capacitat de generalització de l'enfocament a causa de les classes semàntiques relacionades amb

⁸<http://timeml.org/site/timebank/documentation-1.2.html#iaa>

els esdeveniments. Els rols semàntics són útils pel fet que (i) certs rols de verbs específics tenen més probabilitats de contenir esdeveniments, i (ii) els verbs principals, que són normalment els esdeveniments, són identificats per SRL com a nuclis dels predicats. D'altra banda, la semàntica millora especialment el tipus més complex d'esdeveniments: els esdeveniments nominals concentraren el 70% de la millora total.

Classificació d'esdeveniments

Aquesta tasca consisteix en la classificació dels esdeveniments en les classes d'esdeveniments TimeML: *occurrence*, *state*, *aspectual*, *reporting*, *perception*, *i.action*, *i.state*. La taula 4.9 mostra els resultats obtinguts en aquesta tasca en anglès sobre les dades de TempEval-2.

| Approach | partial_accuracy | complete_accuracy |
|-----------------|------------------|-------------------|
| TIPSem-B | 0,79 | 0,80 |
| TIPSem-LS | 0,79 | 0,79 |
| TIPSem-SR | 0,79 | 0,79 |
| TIPSem | 0,78 | 0,78 |

Taula 4.9: Classificació d'esdeveniments (anglès)

En els resultats obtinguts en la classificació d'esdeveniments per a l'anglès, el *baseline* (TIPSem-B) supera lleugerament el nostre enfocament (TIPSem), ja que obté un 0,80 enfront de 0,78. Per tant, la manera en què hem fet servir la semàntica no és prou discriminativa, ja que produeix més errors que millores.

Aquest fet és recolzat per un experiment de validació creuada en el qual s'obté una reducció mitjana d'error nul·la (vegeu l'apèndix C).

- **Contribució de la semàntica lèxica:** Encara que la característica *top4hypers* se suposava que era útil per a la classificació, no hi ha introduït cap millora. Com que fem servir el sentit més freqüent (primer sentit) per a obtenir el valor de *top4hypers*, esdeveniments com *maintain*, que poden representar les diferents classes (p. e.: *occurrence*, *reporting*), sempre obtenen el mateix valor i el model aprèn que la característica no dona informació discriminativa (vegeu l'exemple 105). Com a treball futur, s'aplicaran tècniques de desambiguació del sentit (WSD) per a obtenir la característica *top4hypers* (per exemple, *mantenir-sentit1* i *mantenir-sentit2*) per a millorar la classificació.

(105) TempEval-2 test set - file: WSJ900813-0157 - sentence: 3

- a. The US [maintains *REPORTING*] that under the UN charter, the Kuwaiti request [...]

TempEval-2 train set - file: wsj_0027 - sentence: 1

- b. [...] a satisfactory level of profit is achieved
and [maintained *OCURRENCE*].

- **Contribució dels rols semàntics:** Els esdeveniments verbals com *to complete* poden representar un esdeveniment *occurrence* (p. e.: [*He* _{A0}] *completed* [*the puzzle* _{A1}]) o *aspectual* (p. e.: [*He* _{A0}] *completed* [*the transaction* _{A1}])). No obstant això, ambdues instàncies de *to complete* obtenen els mateixos valors per a les característiques basades en rols. Com a treball futur, s'afegirà una característica que capture si el rol A1 d'un esdeveniment verbal (p. e.: *complete*) conté un altre esdeveniment (p. e.: *transaction*).
- **Errors trobats:** Hem dut a terme un estudi detallat sobre els errors de classificació d'esdeveniments de TIPSem (2010a). La taula 4.10 mostra els resultats per a cada categoria gramatical.

| | | Precision | Recall | F1 |
|----------------|-------------|-----------|--------|-------|
| Classification | <i>Verb</i> | 73,86 | 74,21 | 73,51 |
| | <i>Noun</i> | 62,73 | 41,33 | 49,53 |
| | <i>Adj.</i> | 55,69 | 31,12 | 40,41 |

Taula 4.10: Classificació d'esdeveniments (detall per classe gramatical)

Com es pot observar, la majoria dels errors estan en la classificació d'esdeveniments nominals i adjectivals.

Com es pot veure en els resultats per classe que es mostren a la taula 4.11, algunes classes es classifiquen a molt alt rendiment (*i. e.*, *perception*, *reporting*, *occurrence*), i altres a un menor rendiment (*i. e.*, *aspectual*, *i_action*, *i_state and state*). Aquesta distribució de resultats no coincideixen amb la distribució del exemples d'entrenament de la taula de la formació exemples que es mostren a la taula 4.1. En canvi, els resultats obtinguts assenyalen la complexitat de cada classe diferent. Estem d'acord amb Bethard i Martin (2006) en el fet que els esdeveniments *state* són els més complexos i heterogenis.

Els resultats de classificació d'esdeveniments assenyalen que els trets semàntics proposats necessiten noves millores per a introduir avantatges en

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

| Event class | partial accuracy |
|-------------------|------------------|
| <i>perception</i> | 1,00 |
| <i>reporting</i> | 0,90 |
| <i>occurrence</i> | 0,90 |
| <i>aspectual</i> | 0,66 |
| <i>i_action</i> | 0,56 |
| <i>i_state</i> | 0,43 |
| <i>state</i> | 0,33 |

Taula 4.11: Classificació d'esdeveniments (detall per classe)

la tasca. En l'actualitat, el nostre *baseline* és la millor aproximació per a aquesta tasca.

Processament de relacions temporals

L'avaluació segueix l'estructura de TempEval-2 per mesurar la categorització de relacions temporals. Aquesta tasca es defineix com l'assignació d'una categoria (*i. e.*, *before*, *before-or-overlap*, *overlap*, *overlap-or-after*, *after*, *o vague*) a un parell d'entitats temporals. La tasca de categorització es divideix en quatre subtasques que depenen de les entitats temporals relacionades: categorització de relacions entre esdeveniments i *timex* (event-*timex*), categorització de relacions entre esdeveniments i la DCT (event-DCT), categorització de relacions entre esdeveniments principals (main-event), i categorització de relacions entre esdeveniments subordinats (subord-event). Cada subtasca usa diferents trets semàntics com ara la semàntica lèxica, els rols semàntics, i/o la semàntica temporal.

Categorització de relacions event-timex

Com es descriu en la secció 3.2.3, la tasca de categorització de relacions event-*timex* es defineix com el procediment per a decidir la categoria de relació entre un esdeveniment i una expressió temporal (*timex*) com es mostra en l'exemple (106).

(106) He was <EVENT eid="1">born</EVENT> in <TIMEX3 tid="1">1999</TIMEX3>.

Input: <TLINK lid=1.11"reltype="NONE1.leid="11.1tid="1"/>
Output: <TLINK lid=1.11"reltype="OVERLAP1.leid="11.1tid="1"/>

La taula 4.12 mostra els resultats obtinguts sobre les dades del TempEval-2.

Els resultats mostren una exactitud de 0,68 per l'enfocament TIPSem. L'ús de coneixement semàntic combinat augmenta en 3 punts el *baseline*.

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,65 | 0,65 |
| TIPSem-TS | 0,66 | 0,66 |
| TIPSem-SR | 0,63 | 0,63 |
| TIPSem | 0,68 | 0,68 |

Taula 4.12: Categorització de relacions event-*timex* (anglès)

Individualment, la semàntica temporal augmenta el valor en un punt, però els rols semàntics la fan disminuir en 2. Per tant, les característiques de rols només són útils en combinació amb la semàntica temporal. La combinació fa l'enfocament més potent barrejant la semàntica temporal i les connectives de subordinació temporal.

S'ha provat que la reducció mitjana d'error en una avaluació creuada (5%) és significativa amb una confiança del 90% (vegeu l'apèndix C).

- **Contribució de la semàntica temporal:** L'aplicació de la semàntica temporal (TIPSem-TS) va produir una petita millora en el *baseline* (+0,01). TIPSem-TS té en compte la característica *valdiff*, és a dir, si l'esdeveniment de la relació té a veure amb una *timex*, que no és el de la relació, la característica obté el valor de la diferència temporal (abans, igual, després). Gràcies a aquesta característica, casos com els de (107) que van ser categoritzats erròniament com *overlap* pel *baseline* (TIPSem-B) són correctament categoritzats com a classificats com *after* per TIPSem-TS.

(107) TempEval-2 test set - file: APW19980301.0720 - sentence: 13
a. The Maariv newspaper [said _{e57}] [Sunday _{t1983}] that Netanyahu 's comments [last week _{t1985}] were in response to [...]
(e57 *after* t1985)

TempEval-2 test set - file: wsj_0781 - sentence: 11
b. [...] the chip may have been responsible for a [decline _{e40}] [...] [yesterday _{t283}], and partly responsible for a drop [...] on [Wednesday _{t284}].
(e57 *after* t1985)
(e40 *after* t284)

En l'exemple (107a), la relació entre les entitats subratllades (e57 i t1985) es categoritza correctament com *after* si s'aplica la semàntica temporal, perquè la funció *valdiff* assenyalava que *said* té a veure amb

Sunday (1998-03-01) que és posterior que *last week* (1998-02-22). La mateixa situació ocorre en (107b), on *decline* va passar *yesterday* (1989-10-27), que és, en aquest cas, posterior a *Wednesday* (1989-10-25).

- **Contribució dels rols semàntics:** L'aplicació dels rols semàntics (TIPSem-SR) només millora el *baseline* si s'usa en combinació amb la semàntica temporal. Tenir en compte només les característiques de rols semàntics (*TRoleSBARHead*) produeix errors, ja que, en els casos on es requereixen nivells lingüístics més alts, com la pragmàtica i el coneixement del món real, els rols de forma aïllada no són discriminatòris. Per tant, l'enfocament aprèn un model adhoc (útil només per a les dades d'entrenament). Un exemple es mostra en (108).

(108) TempEval-2 test set - file: APW19980301.0720 - sentence: 0
 a. Top Israeli officials sent strong new signals [*Sunday* _{t123}] that Israel wants to withdraw from southern Lebanon, where a costly [*war* _{e4}] of attrition has been steadily claiming soldiers' lives.

TIPSem-SR categoritza la relació entre *e4* i *t123* com *before*, mentre que l'anotació clau i l'enfocament *baseline* estan d'acord a considerar la relació com a *overlap*. *TRoleSBARHead* en aquest cas val *that*, que no és tan informativa com podria ser *during*, *before* o *when*, per exemple. Aquesta característica va fer que TIPSem-SR canviara la categorització de *overlap* a *before* ja que la majoria de relacions de l'entrenament on *TRoleSBARHead* val *that* (p. e.: *I said yesterday, that the match had been amazing*) són del tipus *before*.

En (108), *war* s'inicia abans que *Sunday*, perquè se superposa i probablement dura més enllà de la referència temporal.

Aquest tipus de relació temporal requereix coneixement sobre el món real (*i. e.*, sobre la guerra de la qual es parla) i també sobre la durada dels esdeveniments comuns (*i. e.*, una guerra normalment dura un cert període de temps) (Pan et al., 2006). Només amb aquest tipus de coneixement l'enfocament pot suposar que la relació discutida en l'exemple és un *overlap*. Això indica que el nostre enfocament no és prou complex per a cobrir aquestes relacions, ja que el coneixement del món i sobre la durada dels esdeveniments està fora de l'abast del sistema de TIPSem.

- **Errors trobats:** Els errors trobats en el millor enfocament (TIPSem) es deuen principalment a una causa: la dificultat de categoritzar les

relacions difuses o febles: *before-and-overlap*, *overlap-or-after*, i *vague*. Alguns exemples es mostren en (109).

- (109) TempEval-2 test set - file: APW19980301.0720 - sentence: 4
 a. Cabinet minister Michael Eitan [said]_{e25} [Sunday _{t129}] that Netanyahu's remarks marked a substantive [change _{e28}] in Israel's position.

TempEval-2 test set - file: wsj.0781 - sentence: 11
 b. Netanyahu's comments [last week _{t1985}] were in response to signals from Syria that it wants to renew the long-stalled peace [talks _{e62}] between the two countries .

En (109a), *change* s'anota com a anterior o superposat a *Sunday* (before-or-overlap), mentre que l'enfocament ha categoritzat la relació com a *overlap*. Si s'analitza la frase, es pot veure que *change* pot haver començat abans que *Sunday*. No obstant això, no està clar si segueix succeint o no en *Sunday*. Aquest tipus de relacions relatives són difícils de classificar, fins i tot per als éssers humans. L'acord entre anotadors és d'al voltant del 80%.

Un altre indicador de la complexitat de la tasca és la categoria *vague*. Aquesta categoria la utilitzen els anotadors humans quan la relació no està clara. Aquest raonament pragmàtic és molt complex per als enfocaments computacionals. En l'exemple (109b), el corpus clau categoritza la relació entre *talks* i *last week* com a *vague* perquè la relació no és clara ni directa. Els errors es deuen a l'escassetat de la categoria *vague* en els exemples d'entrenament i al fet que les característiques no són suficients per a distingir aquest tipus de relacions.

Categorització de relacions event-DCT

Com es descriu en la secció 3.2.3, aquesta tasca consisteix a decidir la categoria de relacions entre un esdeveniment i la data de creació del document (DCT), com es mostra en l'exemple (110).

- (110) DCT: 2001-01-15 (tid="0")
 He was <EVENT eid="1">born</EVENT> in <TIMEX3 tid="1">1999</TIMEX3>.
 Input: <TLINK lid=1.11"reltype="NONE1.leid="11.1tid="0"/>
 Output: <TLINK lid=1.11"reltype="BEFORE1.leid="11.1tid="0"/>

En l'exemple, la categorització indica que *born* (1999) va ser abans del moment en el qual s'escriu el document (2001-01-15).

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,80 | 0,80 |
| TIPSem-TS | 0,82 | 0,82 |
| TIPSem-SR | 0,82 | 0,82 |
| TIPSem | 0,82 | 0,82 |

Taula 4.13: Resultats de categorització de relacions event-DCT (anglès)

La taula 4.13 mostra els resultats obtinguts en les dades del TempEval-2.

En la categorització de relacions event-DCT, la semàntica (TIPSem) millora el *baseline* (TIPSem-B) en 2 punts. Tant la semàntica temporal (TIPSem-TS) com els rols semàntics (TIPSem-SR) van augmentar el rendiment del *baseline*. La combinació (TIPSem) va millorar, però en un sol cas, que no és perceptible en els resultats en una precisió de dos decimals. Això indica que la seua contribució és només en part complementària. Per tant, la utilitat de la combinació és menor en aquest cas.

La reducció mitjana d'error (3%) ha sigut provada mitjançant una validació creuada i la prova t-test mostra que la reducció és significativa amb un 90% de confiança (veure apèndix C).

- **Contribució de la semàntica temporal:** La semàntica temporal ha ajudat en els casos en què existeix una relació sintàctica entre l'esdeveniment de la relació i una *timex*. Per exemple, en (111), on la relació de *declining* amb *first nine months* fa visible que *declining* va ocórrer abans de la DCT.

(111) TempEval-2 test set - file: wsj_520 - sentence: 5

a. But the company has had weak results so far this year, with earnings [declining _{e19}] 43% [...] through the [first nine months _{t100}] of the year.

- **Contribució dels rols semàntics:** Els rols semàntics han ajudat en esdeveniments modificats per partícules temporals com ara l'adverbi *still* en l'exemple (112), el qual assenyala que l'esdeveniment passa després de la DCT.

(112) TempEval-2 test set - file: wsj_0781 - sentence: 28

Compaq, which said it discovered the bugs, still plans to [announce _{e95}] new 486 products.

- **Errors trobats:** El rendiment aconseguit per TIPSem en aquesta tasca és considerablement alt (0,82). Els casos categoritzats erròniament per l'enfocament consisteixen principalment en aquells que representen les relacions difuses o febles (p. e.: *before-or-overlap*, *overlap-or-after*, *vague*). Per exemple, en (113a), *talks* mostra una relació vaga amb la DCT que no és detectada per TIPSem. A més, els casos categoritzats erròniament inclouen errors de *valdiff*, com en (113b), on *1978* només es refereix a *resolution*, però no a *calling*. Finalment, algunes construccions condicionals també han sigut erròniament categoritzades. Per exemple, en (113c), *making* depèn d'un esdeveniment futur condicional, *cooperate*, la qual cosa indica que ocorrerà després de la DCT, i no abans, com TIPSem resol.

(113) TempEval-2 test set - file: APW19980301.0720 - sentence: 14
a. The newspaper said Syria had indicated its interest in resuming
[talks _{e67}] in [...]

TempEval-2 test set - file: APW19980301.0720 - sentence: 3
b. That [1978 _{t128}] resolution [calls _{e16}] for Israel's unconditional withdrawal [...]

TempEval-2 test set - file: APW19980301.0720 - sentence: 2
c. If the Lebanese government will [cooperate _{e10}] with us in [making _{e11}] appropriate [...]

Els errors reportats suggereixen la inclusió de característiques més complexes per a trobar relacions temporals febles, la millora del càlcul de *valdiff* i el tractament de les dependències d'esdeveniments condicionals.

Categorització de relacions main-event

Com es descriu en la secció 3.2.3, aquesta tasca consisteix a decidir la categoria de relació entre dos esdeveniments principals en oracions diferents (interoracional, frases consecutives). Es mostra un exemple en (114).

(114) He was <EVENT eid="1">born</EVENT> in London.
He <EVENT eid="2">obtained </EVENT> his degree in computer science
at Cambridge University. However, he had <EVENT eid="3">studied</EVENT>
previously in Oxford.

Input:
<TLINK lid=1.11"reltype="NONE1.leid="11.leid="2"/>
<TLINK lid=1.12"reltype="NONE1.leid="21.leid="3"/>


```
Output:
<TLINK lid=1·11"reltype="BEFORE1·leid="11·leid="2"/>
<TLINK lid=1·12"reltype="AFTER1·leid="21·leid="3"/>
```

La taula 4.14 mostra els resultats obtinguts en les dades del TempEval-2. Com que l'únic tipus de semàntica utilitzada és la semàntica temporal, TIPSem-TS és equivalent a TIPSem.

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,58 | 0,58 |
| TIPSem | 0,61 | 0,61 |

Taula 4.14: Categorització de relacions main-event (anglès)

TIPSem obté una exactitud de 0,61. En relació amb la categorització d'esdeveniments principals, la contribució de la semàntica es quantifica en 3 punts per sobre del *baseline*. Una reducció mitjana d'error del 3% s'obté mitjançant una validació creuada detallada en l'apèndix refappendix:TIPSem-English-10fold, que assenyala també que la contribució és significativa amb un nivell de confiança del 99,5%.

- **Contribució de la semàntica temporal:** Aquesta va ajudar a la categorització dels esdeveniments mitjançant la interpretació de la seua relació amb les *timex*. Un exemple de dos esdeveniments (e55 i e57) que van ser categoritzats erròniament com a *overlap* pel *baseline* i van ser categoritzats correctament com a *before* per TIPSem es mostra en (115).

(115) TempEval-2 test set - file: APW19980301.0720 - sentence: 12 and 13
 In addition, 73 soldiers were [killed *e55*] en route to Lebanon when two
 military helicopters crashed [last February *t1982*].
 The Maariv newspaper [said *e57*] [Sunday *t1983*] that [...]

En l'exemple, clarament, el fet que e55 estiga relacionat amb *last February*, mentre que e57 apareix amb *Sunday*, i, per tant, el valor de *valdiff* és *before*, fa que l'enfocament TIPSem obtinga la categorització correcta.

- **Errors trobats:** Igual que en els altres tipus de relacions, els errors de categorització de TIPSem es deuen principalment a les relacions febles (p. e.: *vague*), i també a l'obtenció de valors incorrectes de

valdiff. D'altra banda, les relacions interoracionals complicades com la causalitat no es resolten correctament, com es mostra en l'exemple (116).

- (116) a. John pushed Marc. Marc fell.
b. Marc fell. John pushed him.

Els esdeveniments *pushing* i *falling* descrits en (116a) i (116b) obtenen les mateixes característiques, mentre que en el primer cas la relació és *before*, i en el segon és *after*. El sistema classifica ambdues com *before*, ja que no té en compte la relació causal d'aquests esdeveniments i aprèn que els esdeveniments amb el mateix temps verbal generalment es troben ordenats en la narració.

Categorització de relacions subord-event

Com es descriu en la secció 3.2.3, aquesta tasca consisteix a decidir la categoria de relació entre dos esdeveniments en la mateixa frase (intraoracional) on un esdeveniment sintàcticament domina l'altre. En (117) n'apareix un exemple.

(117) We <EVENT eid="1">saw</EVENT> an <EVENT eid="2">explosion</EVENT>.

Input: <TLINK lid=1·11"reltype="NONE1·leid="11·leid="2"/>
Output: <TLINK lid=1·11"reltype="OVERLAP1·leid="11·leid="2"/>

La taula 4.15 mostra els resultats obtinguts en les dades de TempEval-2. Atès que l'única classe de la semàntica utilitzada és la semàntica lèxica, TIPSem-LS és equivalent a TIPSem.

| Approach | partial accuracy | complete accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,61 | 0,61 |
| TIPSem | 0,66 | 0,66 |

Taula 4.15: Categorització de subord-event (anglès)

Els resultats mostren una exactitud de 0,66 mitjançant TIPSem, de manera que millora el *baseline* en 5 punts. La semàntica introdueix una reducció d'error mitjana del 2%, com es detalla en l'apèndix C, on també es demostra que la contribució és significativa amb una confiança del 95%.

- **Contribució de la semàntica lèxica:** La consideració de la classe d'esdeveniments va ajudar, per exemple, en la categorització d'esdeveniments *i_action* relacionats amb esdeveniments *occurrence* perquè la seqüència d'algunes classes (p. e.: *i_action i occurrence*), en algunes combinacions de temps verbals, és més probable que representen relacions temporals específiques (p. e.: *before*). La consideració de la classe d'esdeveniments millora en els casos com els mostrats en (118).

(118) TempEval-2 test set - file: PRI19980303.2000.2550 - sentence: 0
a. Tomorrow the board of supervisors of Loudon county, Virginia, will [vote _{e1}] on whether a school now located in Mount Vernon can [relocate _{e3}] to their county.

TempEval-2 test set - file: wsj_0781 - sentence: 0
b. Intel Corp.'s most powerful computer chip has flaws that could [delay _{e2}] several computer makers' marketing [efforts _{e3}] [...]

En les frases exemple hi ha esdeveniments *i_action* relacionats amb els esdeveniments *occurrence* que marquen una categoria *before* en la subordinació. En (118a), *vote* és anterior a *relocate*. La consideració de la classe és útil per a disminuir la influència del temps verbal que és futur en *vote* i present en *relocate*, cosa que suggereix una relació *overlap-or-after*. En (118b) tenim una situació similar amb *delay* (*i_action*) i *efforts* (*occurrence*), que gràcies a la consideració de la classe es categoritzen com una relació *before*.

- **Errors trobats:** Els principals errors pel que fa a les relacions de subordinació temporal entre esdeveniments són causats per les següents raons. Primeres relacions, borroses o febles, que són inherentment complexes com hem indicat anteriorment. Per exemple, en (119a), la relació entre *wants* i *stay* és anotada en el corpus clau com a *before-or-overlap*, però TIPSem la categoritza com a *before*. En segon lloc, en les relacions que requereixen un coneixement del món real sobre la durada dels esdeveniments o de la causalitat. Per exemple, en (119b), *charged* succeeix després de *bombing*, però TIPSem categoritza la relació com a *overlap*. Per tal de categoritzar correctament aquesta i altres relacions complexes, és necessari conèixer la relació causal entre *bombing* i *charging*. La causalitat indica que el bombardeig ha de passar abans que la condemna.

- (119) TempEval-2 test set - file: APW19980301.0720 - sentence: 6
 a. No one [wants e_{31}] to [stay e_{32}] in Lebanon [...]

- TempEval-2 test set - file: CNN19980227.2130.0067 - sentence: 12
 b. Rudolph has already been [charged e_{13}] in the [bombing e_{14}].

En aquest apartat, hem presentat els resultats del nostre enfocament semàntic i el nostre *baseline* morfosintàctic. En el següent apartat l'exercici de TIPSem s'analitza a través de la seua comparació amb l'estat de la qüestió.

Comparació amb l'estat de la qüestió

Els resultats de TIPSem i els millors resultats obtinguts en la competició TempEval-2 pels sistemes de l'estat de la qüestió (SOA) es resumeixen en una taula comparativa i s'il·lustren en un gràfic. Es mostren els resultats d'F1 per a les tasques de reconeixement, *partial_accuracy(PA)* per a la classificació i normalització, i *complete_accuracy(CA)* per a la categorització de relacions. Tot i que hem explicat abans que *PA* no ens permet comparar els diferents sistemes, hem mantingut aquesta mesura perquè és l'únic resultat que tenim dels enfocaments del SOA (l'únic reportat en l'avaluació TempEval-2).

La taula 4.16 resumeix els resultats d'anglès. D'altra banda, la figura 4.2 il·lustra aquests resultats indicant els de TIPSem en color fosc i els millors resultats del SOA en color clar.

| Element | Task | TIPSem | SoA |
|---------|-------------------------------|-------------|---------------------------|
| TIMEX | recognition F1 | 0,88 | 0,86 (HeidelTime) |
| | classification <i>PA</i> | 0,98 | 0,98 (TERSEO) |
| | normalization <i>PA</i> | 0,84 | 0,85 (HeidelTime) |
| EVENT | recognition F1 | 0,85 | 0,80 (Edinburgh) |
| | classification <i>PA</i> | 0,79 | 0,77 (TRIOS) |
| TLINK | event- <i>timex</i> <i>PA</i> | 0,68 | 0,65 (TRIOS) ^a |
| | event-DCT <i>PA</i> | 0,82 | 0,80 (JU-CSE) |
| | main-events <i>PA</i> | 0,61 | 0,58 (TRIPS) ^b |
| | sub-events <i>PA</i> | 0,66 | 0,66 (NCSU-indi) |

^aExceptionally, TRIOS categorized only the 75% of relations (*PA*)

^bExceptionally, TRIPS categorized only the 83% of relations (*PA*)

Taula 4.16: Resum de resultats per a l'anglès

4.1. AVALUACIÓ SOBRE TEXT EN ANGLÈS

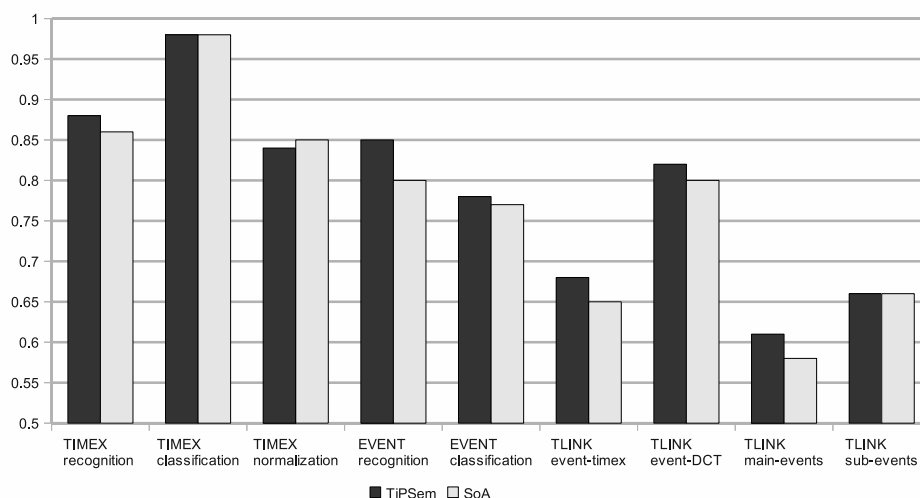


Figura 4.2: Comparativa de resultats amb l'estat de la qüestió (anglès)

TIPSem supera o iguala el rendiment de l'estat de la qüestió en totes les tasques. Les diferents subtasques s'analitzen a continuació.

Pel que fa al processament de *timex*, en el reconeixement, TIPSem (0,88) va superar el SOA (0,86). En la classificació de *timex* i la normalització s'obtenen resultats competitiu. No obstant això, la mètrica *PA* no és comparable com es va discutir en la secció 4.1. Per tant, els valors només donen una estimació indirecta.

Pel que fa al processament d'esdeveniments, en el reconeixement, TIPSem (0,85) va superar el SOA (0,80). En la classificació s'obtenen resultats competitiu, però la mètrica *PA* només dona una estimació indirecta. El rendiment assolit en les tasques de processament d'esdeveniments és una de les majors aportacions d'aquesta tesi. Val la pena esmentar que, en l'avaluació oficial TempEval-2, TIPSem va obtenir les puntuacions més altes per al reconeixement d'esdeveniments i la seua classificació (Llorens et al., 2010b). Es poden trobar més detalls en l'apèndix D.

En matèria de tractament de relacions temporals, TIPSem supera el SOA en totes subtasques de categorització⁹ excepte en les relacions d'esdeveniments subordinats, on ambdós (TIPSem i SOA) obtenen el mateix rendiment (0,66).

En línies generals, els resultats indiquen que l'enfocament que es presenta

⁹En aquestes tasques, *PA* és equivalent a *CA* excepte per al sistema TRIPS/TRIOS.

és molt competent per a abordar la tasca de processament d'informació temporal. TIPSem supera o iguala el rendiment de SOA en totes les subtasques. En particular, en el processament d'esdeveniments, el nostre enfocament representa una millora considerable respecte al SOA, el qual assenyala que l'aproximació semàntica és útil per aconseguir un alt rendiment.

4.2 Avaluació multilingüe: castellà, italià, i xinès

Per a oferir una perspectiva multilingüe del problema abordat en aquest treball, els enfocaments presentats han sigut avaluats en altres idiomes a més de l'anglès. En les avaluacions ens interessa analitzar si la inclusió de coneixement semàntic millora el nostre enfocament d'una manera similar per a diferents idiomes.

Es va dur a terme una extensió completa de l'aproximació a l'espanyol, i també dues extensions parcials a l'italià (reconeixement i classificació d'esdeveniments utilitzant només la semàntica lèxica) i al xinès (reconeixement de *timex* i esdeveniments utilitzant només els rols semàntics), com es va detallar en la secció 3.4.

Avaluació sobre text en castellà: TIPSemES

L'extensió de l'espanyol és la més completa, ja que abarca totes tasques de processament de la informació temporal. Aquesta s'ha avaluat utilitzant el criteri TempEval-2 descrit en la secció 4.1, però amb les dades espanyoles.

Les dades de TempEval-2 per a l'espanyol (Saurí, Saquete, & Pustejovsky, 2009) provenen del corpus Ancora (Taulé et al., 2008). Ancora és el major corpus disponible anotat en diferents nivells lingüístics en espanyol. Es compon de 500K paraules principalment de textos periodístics. El corpus ha sigut anotat i revisat manualment en el nivell morfològic (PoS i lemes), sintàctic (constituents i funcions), i semàntic (rols, entitats nomenades, i sentits de WordNet) (Navarro, Cívít, Martí, Marcos, & Fernández, 2003). Això ens permet obtenir les característiques del TIPSem directament des del corpus. La taula 4.17 resumeix les estadístiques de conjunt de dades.

La taula 4.17 mostra la densitat dels elements TimeML en les dades, així com els tipus més populars i classes. Es pot observar que hi ha molts més esdeveniments (11274) que *timex* (1186), que la classe d'esdeveniment més freqüent és *occurrence* (3628 de 11274), i que el tipus més freqüent de *timex* és *date* (792 de 1186). El tipus més popular de relació és esdeveniment-DCT (11553). Mirant en cada tipus de relació, per a l'esdeveniment-*timex* la categoria més popular és *overlap* (683 de 855), i per l'esdeveniment-DCT

4.2. AVALUACIÓ MULTILINGÜE

| Set | Docs | Words | Elements: classes |
|-------------------------|------|-------|--|
| Training | 175 | 58K | TIMEX(1094) : date(727), duration(274), time(56), set(37) EVENT(10449) : occurrence(3428), state(3281), i_state(1625), i_action(1375), reporting(541), aspectual(153), perception(46) TLINK event-time (772): overlap(616), before(54), after(25), before-or-overlap(16), overlap-or-after(45), vague (16) TLINK event-dct(10445) : overlap(4069), before(3665), after(1601), before-or-overlap(580), overlap-or-after(393), vague (137) |
| Test (entities) | 17 | 4K | TIMEX(92) : date(65), duration(24), time(1), set(2) EVENT(825) : occurrence(200), state(318), i_state(149), i_action(87), reporting (57), aspectual(13), perception(1) |
| Test (relations) | 18 | 6K | TLINK event-time (83): overlap(67), before(8), after(2), before-or-overlap(1), overlap-or-after(2), vague (3) TLINK event-dct(1108) : overlap(545), before(312), after(121), before-or-overlap(67), overlap-or-after(22), vague (41) |

Taula 4.17: Estadístiques de les dades TimeML en castellà (TempEval-2)

és també *overlap* (4614 de 11.553), seguit de *before*. Com es pot observar, la distribució pràcticament coincideix amb la distribució de l'anglès.

Processament de *timex*

Pel que fa al processament de *timex*, l'extensió espanyola cobreix les seues tres subtasques: el reconeixement, classificació i normalització.

Reconeixement de *timex*

La taula 4.18 mostra resultats obtinguts en les dades de prova TempEval-2.

| Approach | precision | recall | F1 |
|---------------|-----------|--------|-------------|
| TIPSem-B | 0,97 | 0,81 | 0,88 |
| TIPSem-LS | 0,96 | 0,85 | 0,90 |
| TIPSem-SR | 0,96 | 0,88 | 0,91 |
| TIPSem | 0,96 | 0,89 | 0,92 |

Taula 4.18: Reconeixement de *timex* (espanyol)

TIPSem va obtenir una F1 molt alta (0,92). En espanyol, l'aplicació

de la semàntica lèxica, els rols semàntics, i la seua combinació van obtenir una reducció de l'error en F1 respecte al *baseline* del 17%, 25%, i el 33% respectivament. Per tant, la combinació d'ambdós tipus de coneixement semàntic obté la major reducció de l'error (33%), la qual cosa indica que la seua contribució és complementària.

La significativitat de la millora de la F1 ha sigut provada amb una validació creuada. Una prova t-test aparellada sobre els resultats suporta que la reducció de l'error és estadísticament significativa, amb un 95% de confiança ¹⁰.

Els trets semàntics condueixen a models més generals. Per exemple, la semàntica millora la delimitació de *timex* llargues (120a). També, els casos ambigus, com en (120b), on *12* i *14 horas* són *timex*, i el *baseline* només va poder anotar *14 horas* com a *timex*, perquè és un nombre seguit d'una unitat temporal. Els rols semàntics ajuden al reconeixement de *12* com a *timex*, ja que és un nombre amb un rol semàntic temporal.

- (120) TempEval-2 test set - file: 108.20000601.a.txt - sentence: 10
 a. El caso, que dura ya más de dos años, [...]
 TempEval-2 test set - file: 10920.20001115.txt - sentence: 8
 b. [...] les obliga a trabajar hasta 12 ó 14 horas [...]

En poques paraules, en espanyol, l'anàlisi condueix a les mateixes conclusions obtingudes per a l'anglès.

Classificació de *timex*

La taula 4.19 mostra els resultats obtinguts en les dades de TempEval-2. Només la semàntica lèxica s'utilitza en aquesta tasca ¹¹, per tant, TIPSem-LS i TIPSem són equivalents.

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,99 | 0,95 |
| TIPSem | 1.00 | 1.00 |

Taula 4.19: Classificació de *timex* (castellà)

Centrant-nos en l'exactitud completa (*complete_accuracy*), que és l'única mesura comparable, el rang de puntuacions varia entre 0,95 (TIPSem-B) a

¹⁰t(9)=1.97 (one-tail $p=0,04$). p : probabilitat d'obtenir la millora per l'atzar.

¹¹Vam comprovar que si els rols semàntics s'hi inclouen, el rendiment del *baseline* no es millora (com en anglès).

1,00 (TIPSem). Aquests resultats indiquen que l'ús de la semàntica lèxica és útil per a aquesta tasca (1,00). En particular, la contribució de la semàntica lèxica se centra en les característiques obtingudes de TIMEK (p. e.: el patró i la granularitat) i no en la característica *top4hypers* extreta de WordNet.

La millora va afectar principalment la durada com *media hora* i *dos años* que van ser classificats incorrectament pel *baseline* com a *date*, però correctament classificades per TIPSem com a *duration* gràcies al valor patró (*i. e.*, Num_TUnit).

Durant l'experimentació s'ha trobat una inconsistència entre l'anglès i l'espanyol. En espanyol, l'expressió *hace dos años* i totes les expressions que segueixen el patró *hace_Num_TUnit* van ser anotades com a *duration*, mentre que en anglès que van ser anotades com a *date*.

En resum, els resultats obtinguts per espanyols confirmen les conclusions de l'anglès: les característiques basades en la semàntica lèxica són discriminatòries, mentre que els rols semàntics no ajuden en aquesta tasca.

Normalització de *timex*

La taula 4.20 mostra els resultats obtinguts en les dades de TempEval-2. TIPSem-LS i TIPSem SR no es mostren en la taula perquè en la normalització els rols semàntics no s'utilitzen, i, per tant, TIPSem-LS és equivalent a TIPSem.

| Approach | step 1 | step 2 | partial_accuracy | complete_accuracy |
|---------------|--------|--------|------------------|-------------------|
| TIPSem-B | 0,73 | - | 0,86 | 0,78 |
| TIPSem | 0,89 | 0,89 | 0,90 | 0,87 |

Taula 4.20: Normalització de *timex* (castellà)

En termes generals, es pot observar que TIPSem va obtenir un rendiment de 0,87, que millora el *baseline* en +0,09%. Això es deu a la contribució de la semàntica lèxica en el primer pas de l'estratègia de normalització de TIPSem.

En el primer pas, TIPSem-B obté un *complete_accuracy* de 0,73 i TIPSem de 0,89. Igual que en anglès, la millora obtinguda per TIPSem es deu a la semàntica lèxica obtinguda de TIMEK (p. e.: *pattern*).

En el segon pas, els errors detectats són de dos tipus: (i) errors humans en l'anotació de les dades, i (ii) errors de TIPSem en elisions i referències anafòriques. Alguns exemples apareixen en (121).

- (121) TempEval-2 test set - file: 108_20000301.c.txt - sentence: 3
 a. [...] su regreso a las pistas después de un largo periodo [...]

CAPÍTOL 4. AVALUACIÓ

TempEval-2 test set - file: 108_20020202.txt - sentence: 1

b. [...] la federación se pronunciará ese día sobre los casos de dopaje [...]

TempEval-2 test set - file: 108_20020202.txt - sentence: 2

c. La federación suspendió con cuatro meses a Guardiola y con cinco a Stam.

En (121a), TIPSem normalitza l'expressió d'un període indefinit (PXX), però en l'anotació clau es va normalitzar en un període indefinit de mesos (PXM). Des del nostre punt de vista, els períodes indefinits haurien de normalitzar-se com a PXX a menys que tinguem una unitat temporal que participe en l'expressió (p. e.: many months, some months). En les dades d'anglès, aquests casos van ser anotats d'aquesta manera, per la qual cosa suposem que aquests errors constitueixen un problema de l'anotació manual de les dades en espanyol.

No obstant això, en (121b) i (121c), ens trobem amb un problema del sistema que també es va observar en les dades d'anglès. Aquest es relaciona amb la complexitat d'algunes expressions temporals. En el primer cas, el text està incomplet i la referència real de *that day* no es pot trobar en el text, de manera que aquesta *time*x no es pot normalitzar. En el segon cas, per tal de normalitzar *five* el sistema necessita resoldre la elisió de *months* que requereix resoldre la relació de coreferència amb *four months*.

En resum, la normalització d'expressions temporals en espanyol pot ser resolta computacionalment a un rendiment acceptable, per sobre de 0,80, però superar aquesta barrera és complicat, perquè es necessiten capacitats de resolució de l'anàfora i processament del discurs, com ja hem comentat per a l'anglès.

Processament d'esdeveniments

Aquesta tasca inclou el reconeixement i la classificació d'esdeveniments.

Reconeixement d'esdeveniments

La taula 4.21 mostra els resultats obtinguts sobre les dades de TempEval-2.

Els resultats del reconeixement d'esdeveniments per TIPSemES són alts (0,90) i, com en anglès, la combinació d'ambdós tipus de coneixement semàntic obté les puntuacions més altes, la qual cosa indica que les dues són complementàries.

La significança de la millora en F1 de la semàntica ha sigut provada amb una validació creuada. Una prova t-test aparellada sobre els resultats

| Approach | precision | recall | F1 |
|---------------|-----------|--------|-------------|
| TIPSem-B | 0,90 | 0,86 | 0,88 |
| TIPSem-LS | 0,93 | 0,87 | 0,90 |
| TIPSem-SR | 0,92 | 0,87 | 0,89 |
| TIPSem | 0,92 | 0,88 | 0,90 |

Taula 4.21: Reconeixement d'esdeveniments (castellà)

suporta que la reducció de l'error és estadísticament significativa amb un 99% de confiança ¹².

La semàntica ha contribuït a aprendre models més generals, el fet que millora el reconeixement dels esdeveniments nominals. L'exemple (122) mostra un cas en què tots dos tipus de semàntica han contribuït al reconeixement correcte.

(122) TempEval-2 testset - file: 108.20000301.b.txt - sentence: 8
Valero Ribera no disimuló su preocupación por el sorteo y [...]

En (122), el nom *sorteo* és un esdeveniment que no va reconèixer el *baseline*. No obstant això, la combinació de la semàntica lèxica (*top4hypers* = *activity* = *event* = *abstraction*) i els rols semàntics (*role* = *A1*) va ajudar al reconeixement correcte.

En resum, el rendiment del reconeixement d'esdeveniments en espanyol és alt (0,90). Això assenyalava que, igual que per a l'anglès, els trets semàntics de TIPSem són útils per a aquesta tasca.

Classificació d'esdeveniments

La taula 4.22 mostra els resultats obtinguts sobre les dades de TempEval-2.

| Approach | partial_accuracy | complete_accuracy |
|---------------|------------------|-------------------|
| TIPSem-B | 0,66 | 0,66 |
| TIPSem-LS | 0,66 | 0,67 |
| TIPSem-SR | 0,66 | 0,67 |
| TIPSem | 0,66 | 0,68 |

Taula 4.22: Classificació d'esdeveniments (castellà)

L'aplicació de les característiques semàntiques de classificació d'esdeveniments incrementa els resultats de 0,66 a 0,68. La principal causa de la mi-

¹²t(9)=3.16 (one-tail $p=0,006$). p : probabilitat d'obtenir la millora per l'atzar.

llora obtinguda és la correcta classificació dels esdeveniments nominals com a *occurrence*. Alguns esdeveniments nominals com *oferta*, *salto*, *segregación*, que no estan en les dades d'entrenament, són classificats incorrectament pel *baseline* amb la classe més freqüent (*state*).

Pel que fa als resultats obtinguts per a cada classe, novament *perception* (1,0), *reporting* (0,86) i *occurrence* (0,77) han obtingut els resultats més alts, mentre que *i_state* (0,55), *aspectual* (0,54) and *i_action* (0,46) han obtingut un menor rendiment. No obstant això, *state* ha obtingut una puntuació de 0,72.

Encara que les bases de dades d'anglès i espanyol són diferents en contingut i els resultats no són directament comparables, s'han construït a partir d'articles de premsa i una puntuació de 0,72 en la classificació de *state* en espanyol és molt més alta que la puntuació obtinguda en l'experiment en anglès (0,33). A més, la proporció d'esdeveniments *state* en espanyol és molt més alta que en anglès. S'han analitzat dos conjunts de dades i s'han observat que aquesta diferència es produeix per una inconsistència d'anotació en el conjunt de dades espanyol: totes les instàncies del verb *to be* estan anotades com a *state*, cosa que no passa en el conjunt de dades anglès i que no queda molt ben definit a les directrius TimeML.

Processament de relacions temporals

En espanyol, la tasca de categorització s'ha limitat a dues subtasques: la categorització de relacions event-*timex* i event-DCT.

Categorització de relacions event-*timex*

La taula 4.23 mostra els resultats obtinguts en les dades de prova TempEval-2.

| Approach | complete_accuracy |
|-----------|-------------------|
| TIPSem-B | 0,81 |
| TIPSem-TS | 0,81 |
| TIPSem-SR | 0,81 |
| TIPSem | 0,81 |

Taula 4.23: Categorització de relacions event-*timex* (castellà)

L'exactitud obtinguda per TIPSem és alta (0,81), però l'ús de la semàntica no produeix cap millora sobre el *baseline*.

Pel que es refereix a la distribució de les categories en les relacions event-*timex* en espanyol, es pot observar que el 80% (616/772) de les relacions es

classifiquen com a *overlap*. Tenint en compte que TIPSem està principalment basat en aprenentatge automàtic i utilitza la informació estadística per a aprendre els models de categorització, el fet anterior fa que la categorització de TIPSem tendisca a ser per defecte *overlap*. Mirant les dades de test, s'observa que el 81% (67/83) de les relacions són *overlap*. Finalment, mirant l'eixida de TIPSem en aquesta tasca per a l'espanyol, es pot observar que anota la majoria de les categories com a *overlap* (totes menys dues anotades com a *overlap-or-after*). Això explica el rendiment obtingut i la nul·la millora observada quan s'utilitza la semàntica: TIPSem ha après un bon *baseline* per a l'espanyol (anota gairebé totes les relacions com a *overlap*), i els trets semàntics no són suficientment discriminatoris per modificar el *baseline* après.

Categorització de relacions Event-DCT

Aquesta tasca tracta de decidir la categoria de les relacions entre un esdeveniment i la data de la creació dels documents (DCT). La taula 4.24 mostra els resultats obtinguts en les dades de TempEval-2.

| Approach | complete_accuracy |
|-----------|-------------------|
| TIPSem-B | 0,59 |
| TIPSem-TS | 0,59 |
| TIPSem-SR | 0,59 |
| TIPSem | 0,59 |

Taula 4.24: Categorització de relacions Event-DCT (castellà)

L'exactitud obtinguda per TIPSemES en les relacions event-DCT és de 0,59, i, segons sembla, com en les relacions *timex-event*, l'ús de la semàntica no produeix cap millora pel que fa al *baseline*. No obstant això, si els resultats s'analitzen en detall, TIPSem (655/1108 - 59,1%) millora lleugerament TIPSem-B (652/1108 - 58,8%). Però això és inapreciable en les puntuacions amb dos decimals.

Observant les diferències entre el *baseline* i el nostre enfocament, es troba que tot i que TIPSem millora en 62 casos incorrectament classificats pel *baseline*, també empitjora el *baseline* en 59 casos.

La millora i, en general, l'efecte de la semàntica se centra en la semàntica temporal (TIPSem-TS) i no en els rols semàntics (TIPSem-SR).

Els casos en què l'ús de la semàntica temporal condueixen a la correcta categorització es deuen a la característica *valdiff*. L'exemple que es mostra en (123) posa de manifest que no hi ha cap problema per al *baseline* en la

categorització d'esdeveniments verbals amb la DCT com *salió*, on el temps verbal passat és prou fort per a assenyalar que l'esdeveniment va ocórrer abans que la DCT. No obstant això, en el cas dels esdeveniments nominals, on no hi ha temps verbal, com en *defensa*, *valdiff* és l'únic indicador per a la categorització de la relació.

- (123) TempEval-2 test set - file: 108.19990501.txt - sentence: 1
El Gobierno Chileno [*salió* _{e1}] [*ayer* _{t1}] en [*defensa* _{e2}] de [...]

En l'exemple, *defensa* està sintàcticament relacionat amb *ayer*, la qual cosa indica que la relació és *before*.

No obstant això, en altres casos, la consideració de *valdiff* condueix a la categorització errònia. Això es deu al fet que *valdiff* no es va calcular amb la *timex* adequada com a (124).

- (124) TempEval-2 test set - file: 108.20000301.a.txt - sentence: 16
Los corredores europeos se [*quejaron* _{e93}] de la poca [*presencia* _{e94}] policial en la [*prueba* _{e95}] que [*empezó* _{e96}] [el 25 de febrero _{t17}] y [*concluye* _{e97}] [el día 5 _{t18}].

En aquest exemple, a l'hora de categoritzar *presencia* la *valdiff* es va calcular amb *el día 5* i, per tant, va obtenir una categoria *after*. No obstant això, el sentit comú dicta que *poca presencia* va succeir abans que la DCT – probablementment durant alguns dies després del començament de la prova (febrer 25). A causa d'aquests errors la influència de la semàntica resulta imperceptible. Són necessàries altres millores per a superar el *baseline*.

Alguns dels errors encara sense resoldre, tant pel *baseline* com per TIPSem, com en anglès, són causats per les relacions difuses o febles, com en (125).

- (125) TempEval-2 test set - file: 108.19990501.txt - sentence: 9
Josep Piqué , [*manifestó* _{e43}] [*ayer* _{t6}] que en su [*opinión* _{e44}] [...]

En aquest exemple, *opinión* és vàlida abans de la DCT (*ayer*), però també pot ser vàlida en el moment de l'escriptura (DCT). En l'anotació clau, la relació és la categoria *overlap*, mentre que TIPSem categoritza la relació com a *before*.

En aquest apartat, hem presentat els resultats del nostre enfocament semàntic per a l'espanyol. En el següent apartat analitzem TIPSemES a través de la seua comparació amb l'estat de la qüestió.

Comparació amb l'estat de la qüestió

De manera anàloga a l'anglès, en aquesta subsecció es resumeixen els resultats prèviament analitzats per a l'espanyol en un quadre comparatiu i la seua taula associada. Això inclou els resultats de nostre enfocament (TIPSem) i el millor estat de la qüestió (SOA) presentat a TempEval-2. Es mostren els resultats d'F1 per a tasques de reconeixement, *partial_accuracy(PA)* per a la classificació i normalització, i *complete_accuracy(CA)* per a la categorització. Tot i que hem explicat anteriorment que *PA* no ens permet comparar els diferents sistemes, hem mantingut aquesta mesura perquè és l'únic resultat que tenim dels enfocaments del SOA.

La taula 4.25 resumeix els resultats per a l'espanyol, i la figura 4.3 il·lustra aquests resultats, on apareix TIPSem en color fosc, i els millors resultats del SOA, en color clar.

| Element | Task | TIPSem | SoA |
|---------|-------------------------------|-------------|-------------|
| TIMEX | recognition F1 | 0,92 | 0,88 (UC3M) |
| | classification <i>PA</i> | 1.00 | 0,91 (UC3M) |
| | normalization <i>PA</i> | 0,90 | 0,83 (UC3M) |
| EVENT | recognition F1 | 0,90 | - |
| | classification <i>PA</i> | 0,68 | - |
| TLINK | event- <i>timex</i> <i>CA</i> | 0,81 | - |
| | event-DCT <i>CA</i> | 0,59 | - |

Taula 4.25: Resum de resultats per al castellà

Per a l'espanyol, l'estat de la qüestió es limita al tractament de *timex*. L'enfocament TIPSem supera el rendiment de l'estat de la qüestió en totes les tasques de processament de *timex*, com es mostra en la taula i a la figura (reconeixement +0,04%, classificació +0,09%, i normalització +0,07%). No obstant això, cal destacar que en la classificació de *timex* i la normalització, la comparació només és indirecta, a causa dels problemes discutits en la secció 4.1.

Per al processament d'esdeveniments i relacions temporals, fins on nosaltres sabem, TIPSem és una aproximació pionera en espanyol.

Els resultats per a anglès i espanyol no es poden comparar, perquè encara que tots dos conjunts de dades procedeixen d'articles de notícies, no són paral·lels¹³. Per tant, el fet que els resultats obtinguts són lleugerament superiors per a l'espanyol que per a l'anglès no es pot avaluar. D'alguna manera, dos factors podrien haver influït en aquest fet: (i) l'enfocament

¹³Els uns no són la traducció dels altres.

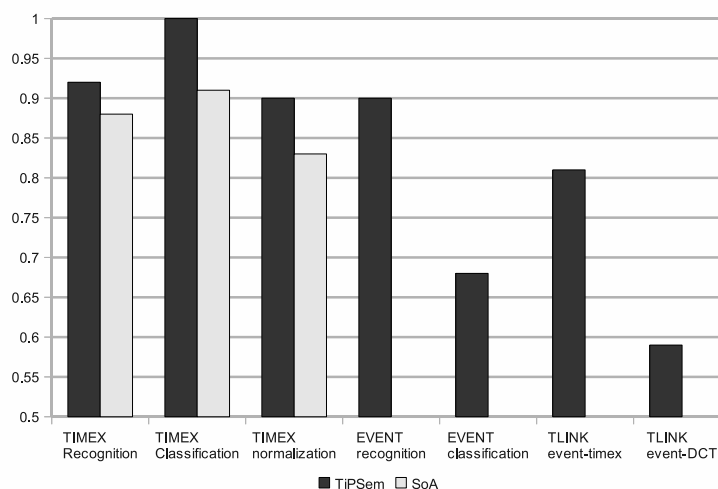


Figura 4.3: Comparativa amb l'estat de la qüestió (castellà)

anglès utilitza eines automàtiques per obtenir les característiques, mentre que l'espanyol les extrau d'un corpus revisat manualment, i (ii) la mida de les dades d'entrenament per a l'espanyol és una mica més gran que en anglès.

En resum, els resultats indiquen que l'enfocament que es presenta és molt competent per a abordar la tasca de processament d'informació temporal. Es pot observar que, igual que per a l'anglès, l'aplicació dels trets semàntics millora el rendiment per a l'espanyol.

Avaluació sobre text en italià: TIPSemIT

Com es descriu en la secció 3.4, TIPSem es va estendre a l'italià per a estudiar el comportament dels trets de la semàntica lèxica en el reconeixement d'esdeveniments i la seua classificació en aquest idioma, i comparar el rendiment amb una aproximació existent basada en regles (TULE converter) (Robaldo et al., 2011).

Es va aplicar el criteri, les mesures i les dades del TempEval-2 per a l'italià. Les dades contenen 4.543 esdeveniments en el conjunt d'entrenament, i 834 en el conjunt de proves. En la taula 4.26 es presenta la distribució dels esdeveniments per classes TimeML de l'entrenament i el test.

Hem establert com a *baseline* per a l'avaluació un sistema basat en regles per a l'italià, el TimeML TULE converter (Robaldo et al., 2011).

| TimeML Classes | # training set | # test set |
|----------------|----------------|------------|
| OCCURRENCE | 2,360 | 456 |
| STATE | 1,591 | 254 |
| LACTION | 288 | 58 |
| LSTATE | 502 | 88 |
| REPORTING | 75 | 18 |
| PERCEPTION | 13 | 1 |
| ASPECTUAL | 75 | 18 |
| Total events | 4,543 | 834 |

Taula 4.26: Dades en italià (esdeveniments TempEval-2)

Avaluem quatre models basats en aprenentatge automàtic de manera incremental afegint característiques. El model bàsic, TIPSemIT_basic utilitza només les característiques bàsiques morfològiques, és a dir, lema, paraula i PoS sense context. Els altres tres models afegeixen al bàsic una millor granularitat de PoS, la combinació de característiques morfològiques i la presència de trets semàntics. En particular, TIPSemIT_FP afegeix el PoS de granularitat fina. TIPSemIT_FPC5 afegeix un context mida de la finestra de cinc amb bigrames de lema i PoS. Finalment, TIPSemIT_FPC5Sem afegeix trets semàntics amb el model anterior.

Reconeixement d'Esdeveniments

La taula 4.27 informa dels resultats per al reconeixement de l'esdeveniment.

| Models | P | R | F1 |
|-------------------------|------|------|-------------|
| TULE converter | 0,84 | 0,74 | 0,79 |
| TIPSemIT_basic | 0,91 | 0,75 | 0,82 |
| TIPSemIT_FP | 0,90 | 0,77 | 0,83 |
| TIPSemIT_FPC5 | 0,89 | 0,81 | 0,85 |
| TIPSemIT_FPC5Sem | 0,91 | 0,83 | 0,87 |

Taula 4.27: Reconeixement d'esdeveniments (italià)

Tot i que tenim un corpus molt reduït, TIPSemIT_basic obté un resultat superior al *baseline* en termes de precisió (0,91 *vs.* 0,84). No obstant això, la cobertura només és millorada en +0,01. Només un reduït nombre d'esdeveniments és reconegut més enllà del *baseline* (625/834 *vs.* 624/834), deixant de banda una gran part dels esdeveniments més difícils, com ara esdeveniments realitzats per substantius, adjectius i sintagmes preposicionals. TIPSemIT_basic pateix d'escassetat de dades. Les reduïdes dimensions del

conjunt d'entrenament i les característiques de baix nivell que s'hi han utilitzat no són suficients per a identificar instàncies no vistes en l'entrenament, ni generalitzar la informació sobre els contextos lingüístics d'ocurrència.

TIPSemIT_FP millora una mica la cobertura i, per tant, també la F1 (0,77 i 0,83), però disminueix la precisió (0,90). El nombre d'esdeveniments reconeguts correctament és superior, però encara per sota de 0,80. L'ús de PoS-detallat és beneficiós per a la detecció dels esdeveniments realitzats per verbs (12 casos), substantius (14 casos) i adjectius (5 casos). Un major nombre de verbs (12) és reconegut causa de la diferència entre els verbs auxiliars i principals. Una anàlisi detallada dels lemes verbals mostra que la millora en el reconeixement es refereix als verbs italians *essere* [ser] i *venire* [venir], que poden ser tant auxiliars com verbs principals. També és interessant notar que alguns verbs realitzats per formes de participi fallen. Això es deu al fet que el POS considera les formes de participi passat com adjectius, quan no són acompanyades per complements específic (p. e.: *PP_da + NP*), la qual cosa fa que identificar-los siga difícil.

Pel que fa a TIPSemIT_FPC5, la precisió és lleugerament inferior als dos models anteriors (0,89), però el model està equilibrat (cobertura = 0,81). L'augment de la cobertura és de +0,07% sobre el *baseline* i +0,06% sobre TIPSemIT_basic. La combinació de PoS resulta una bona estratègia per al reconeixement de molts esdeveniments no verbals (26 substantius, 15 adjectius, i 5 sintagmes preposicionals).

També és interessant notar que aquest model pot detectar els casos d'esdeveniments que estan fora de l'abast del TULE, en particular, els substantius. Per exemple, el nom *fuga* [fuga] en l'exemple 126 és ignorat pel TULE perquè les regles no identifiquen la construcció causativa realitzada per la presència de la preposició *per* [per]. Per contra, TIPSemIT_FPC5 és capaç de superar aquesta limitació, gràcies a la combinació del PoS detallat, i el context.

(126) [...] *evacuta per una fuga di gas*. [evacuat degut a la fuga de gas.]

Finalment, TIPSemIT_FPC5Sem mostra la major cobertura (689/834 = 83%). L'ús del lèxic d'esdeveniments sembla útil per al reconeixement dels esdeveniments nominals (36 casos) i adjectivals (13 casos). L'anàlisi dels lemes substantius que s'han identificat correctament mostra que la majoria d'aquests (52%) són instàncies de tipus punt ¹⁴ (Pustejovsky, 1995),

¹⁴Una entitat complexa que es compon d'almenys dos tipus semàntics diferents que depenen del context (p. e.: *dinar* → *esdeveniment* or *menjar*)

que alternen entre una lectura eventiva i no eventiva com *ricerca* [recerca] o *allargamento* [increment]. La combinació de característiques morfològiques i semàntiques redueix els riscos d'usar aquest recurs per al reconeixement d'esdeveniments. Com es va assenyalar anteriorment, el lèxic d'esdeveniments derivat només inclou els esdeveniments de lectura eventiva de les unitats lèxiques com ara verbs, adjectius i substantius sense desambiguació del sentit. Això significa que les paraules ambigües, és a dir, noms com *assemblea* [meeting/assembly], estan presents en el lèxic només amb la seua lectura eventiva. Encara que la cobertura es veu afectada positivament per la presència dels trets semàntics, la precisió no millora.

Un dels principals impactes del lèxic d'esdeveniments és la reducció de la dispersió de dades. El corpus d'entrenament és petit i per tal d'obtenir generalitzacions sobre les lectures eventives de les unitats lèxiques com substantius i adjectius cal un nombre rellevant de casos. La presència del lèxic supera aquesta limitació.

Classificació d'Esdeveniments

Els enfocaments de classificació han sigut avaluats sobre els esdeveniments reconeguts pel millor model de reconeixement (*i. e.*, TIPSemIT_FPC5Sem). La taula 4.28 mostra els resultats obtinguts.

| Models | Accuracy |
|-------------------------|-------------|
| TULE converter | 0,65 |
| TIPSemIT_basic | 0,74 |
| TIPSemIT_FP | 0,74 |
| TIPSemIT_FPC5 | 0,74 |
| TIPSemIT_FPC5Sem | 0,77 |

Taula 4.28: Classificació d'esdeveniments (italià)

Per a la classificació d'esdeveniments, TULE explota el lèxic d'esdeveniments derivats per accedir a la classe TimeML per defecte i després s'integra aquesta informació amb informació sintàctica. La precisió de TULE és menor (-12 punts) que la obtinguda per TIPSemIT_FPC5Sem. La principal font d'errors de TULE són els errors d'anàlisi que impedeixen l'activació de la regla corresponent, disminuint el nombre d'esdeveniments classificats correctament. La millora del rendiment obtinguda per TIPSemIT_FPC5Sem pel que fa als altres models es deguda a la contribució del lèxic d'esdeveniments. En particular, es registra una millora en la classificació de les classes menys freqüents en les dades, com ara LSTATE (52 % *vs.* 73 %), ASPECTUAL (41% *vs.* 65%) i REPORINT (53 % *vs.* 68 %), amb l'excepció de

L ACTION.

Pel que fa als esdeveniments classificats correctament, l'accés al lèxic d'esdeveniments (TIPSemIT_FPC5Sem) millora la classificació de 50 casos en TIPSemIT_basic, i de 42 casos pel que fa a TIPSemIT_FP i TIPSemIT_FPC5. Val la pena notar que tots dos models (TIPSemIT_FP i TIPSemIT_FPC5) milloren la classificació de TIPSemIT_basic dels verbs en passat participi incorrectament classificats com a OCCURRENCE. Aquesta millora s'esperava, ja que amb el PoS detallat es pot distingir entre les formes del verb finit i no finit. D'altra banda, les finestres de context que diferencien TIPSemIT_FPC5 de TIPSemIT_FP no contribueixen en absolut a millorar la classificació.

Una anàlisi detallada de les classes d'esdeveniments mostra que el tret semàntic (classe per defecte) és clarament un avantatge per a reduir l'impacte de l'escassetat de dades. Per exemple, el verb *simboleggiare* (*simbolitzar*) que no apareix en l'entrenament, es classifica correctament com a STATE només per TIPSemIT_FPC5Sem. Un altre cas en què el lèxic d'esdeveniments demostra la seua validesa es el verb *rafforzare* (*reforçar*). Aquest verb apareix dues vegades en el conjunt d'entrenament, i cada cas pertany a una classe diferent, OCCURRENCE i LSTATE. En el conjunt de test, aquest verb apareix dues vegades, una vegada classificat com a STATE, ja que és realitzat per una forma del participi passat, i l'altra com a LSTATE. TIPSemIT_FP i TIPSemIT_FPC5 poden classificar l'STATE gràcies a la informació del PoS detallat, però no l'LSTATE. Per contra, TIPSemIT_FPC5Sem classifica correctament ambdós casos, i la classificació correcta de l'LSTATE és gràcies al lèxic d'esdeveniments.

No obstant això, el lèxic d'esdeveniments no és perfecte. En particular, hem observat que la cobertura del lèxic, és a dir, el nombre d'entrades, s'ha d'ampliar, especialment per als substantius i adjectius.

Avaluació sobre text en xinès: TIRSemZH

L'extensió TIPSem al xinès (Llorens, Saquete, Navarro, Li, & He, 2011c) ha sigut avaluada sobre les dades del xinès de TempEval-2. Els objectius d'aquesta avaluació són: (i) mesurar el rendiment de TIPSem en xinès, (ii) analitzar l'efecte de l'aplicació dels rols semàntics, i (iii) comprovar el grau d'independència del llenguatge de l'enfocament fora de les llengües indoeuropees.

La taula 4.29 mostra la descripció del conjunt de dades TempEval-2 per al xinès. En l'avaluació, s'ha emprat el criteri de TempEval-2.

Les dades del xinès provenen d'articles del diari *Xinhua News Agency*,

| Set | Docs | Words | Element |
|----------|------|-------|---------------------------|
| Training | 44 | 23K | TIMEX(766) EVENT(3744) |
| Test | 8 | 5K | TIMEX(129) EVENT(1039) |

Taula 4.29: Estadístiques de les dades en xinès (TempEval-2)

excepte els arxius del `chtb.0592` al `chtb.0596`, que provenen d'un article de *Sinorama*, una revista mensual publicada a Taiwan.

En comparació amb l'anglès, les dades d'entrenament són aproximadament la meitat, mentre que les dades de test són de la mateixa mida.

Les subseccions següents informen sobre els resultats obtinguts en aquestes dades.

Reconeixement de *timex*

La taula 4.30 mostra els resultats obtinguts en el reconeixement de *timex* en xinès.

| Approach | precision | recall | F1 |
|-----------------|-----------|--------|-------------|
| TIRSemZH-B | 0,94 | 0,74 | 0,83 |
| TIRSemZH | 0,97 | 0,76 | 0,85 |

Taula 4.30: Reconeixement de *timex* (xinès)

Com es mostra en la taula, TIRSemZH ofereix un alt rendiment, 0,85 F1. L'aplicació dels rols semàntics ha millorat el rendiment del *baseline* en precisió i cobertura, i, per tant, en F1 de 0,83 a 0,85. La reducció d'errors en relació al *baseline* és de 12% en F1. Això recolza la hipòtesi que els rols semàntics ajuden en l'aprenentatge de models més generals, perquè l'enfocament no se centra només en les paraules i PoS, sinó també en els rols que les paraules tenen en els predicats. Això és especialment beneficiós quan la mida de les dades d'entrenament és petita.

En les dades d'entrenament, les *timex* es troben principalment en els arguments etiquetats amb el rol semàntic temporal (AM-TMP). Aquest rol presenta dos avantatges en el reconeixement de *timex*: (i) l'augment de la probabilitat de representar una *timex* de qualsevol paraula que apareix dins d'aquest rol, i (ii) ajudar a l'enfocament en el reconeixement d'expressions temporals multiparaula com mostra 一月底 (final de gener) i 去年六月底 (final de juny de l'any passat), que estan representats per una seqüència de

paraules que tenen un rol AM-TMP. L'exemple (127) il·lustra una *timex* que falla el *baseline* i es reconeguda amb l'ús dels rols semàntics.

- (127) TempEval-2, file: chtb.0628 - sent.: 12
 一季度末，国家外汇储备余额达一千一百二十点六亿美元。
 Al final del primer trimestre, el balanç del canvi de les reserves forasteres va arribar a 112.06 billons de dòlars americans.

En aquesta frase, totes les paraules de l'argument 一季度末 (*final del primer trimestre*) juga el rol AM-TMP. El *baseline* va ignorar aquesta expressió ja que aquesta seqüència específica de paraules no és molt comuna entre els exemples d'entrenament. No obstant això, amb l'ús de rols semàntics, l'enfocament téen compte que juga un rol AM-TMP i, per tant, és probable que siga una *timex*.

En alguns casos, com en (128), l'aplicació dels rols semàntics en va disminuir el rendiment.

- (128) TempEval-2, file: chtb.0593 - sent.: 8
 南韩目前利率已飙至百分之二十。
 L'*actual* interès de Corea del Sud s'ha disparat un vint per cent.

En aquest cas, *interès actual* no té un rol AM-TMP i, per tant *actual* no va ser marcada per l'aproximació com a *timex*. En aquesta expressió, *actual* es un modificador del substantiu *interès*. Com vam discutir en anglès, els modificadors de nom representen un cas complex.

Reconeixement d'Esdeveniments

La taula 4.31 mostra els resultats obtinguts en el reconeixement d'esdeveniments per al xinès.

| Approach | precision | recall | F1 |
|-----------------|-----------|--------|-------------|
| TIRSemZH-B | 0,86 | 0,80 | 0,83 |
| TIRSemZH | 0,90 | 0,80 | 0,85 |

Taula 4.31: Reconeixement d'esdeveniments (xinès)

Com es mostra en la taula, TIRSemZH obté resultats satisfactoris per al reconeixement d'esdeveniments. Els resultats obtinguts mostren que l'aplicació dels rols semàntics augmenten la F1 de 0,83 a 0,85. En el reconeixement d'esdeveniments, la millora se centra en la precisió (29% de reducció d'errors).

A continuació discutim alguns casos representatius en què els rols semàntics han afectat els resultats.

En primer lloc, els rols semàntics han ajudat a augmentar la precisió, és a dir, redueixen els falsos positius que presenta el *baseline*. Hi ha moltes paraules que tenen significats diferents en diferents entorns semàntics. Una paraula *W* pot indicar un esdeveniment en la frase *A*, o simplement una paraula normal en la frase *B*. Mentre que el *baseline* obté falsos positius, l'enfocament semàntic, TIRSemZH, obté el reconeixement correcte.

L'exemple (129) mostra dues frases en les quals apareix la paraula 出. En la primera frase, es tracta d'un verb principal que regeix dos arguments: *Annan*(A0) i *diners* (A1). En aquest cas, 出 (*paga*) és un esdeveniment. En la segona frase, 显示 (*mostrant*) és el verb principal i 出 (*hi*) no és un esdeveniment sinó un modificador verbal. Per tant, 出 té diferents significats en diferents entorns semàntics i, per tant, els rols semàntics ajuden a determinar en quins casos es tracta d'un esdeveniment.

(129) TempEval-2, file: chtb.0595 - sent.: 29

安南不出钱。

Annan no paga els diners.

TempEval-2, file: chtb.0604 - sent.: 11

显示出成功率高、经营状况良好的特点。

Mostrant-hi un gran èxit i les característiques d'un bon negoci.

En l'exemple (130), el *baseline* reconeix 投资 (*inversió* com a esdeveniment. Això deu al fet que aquesta paraula representa un esdeveniment en molts exemples de les dades d'entrenament. De vegades *inversió* representa un esdeveniment. No obstant això, no sempre és així. En la frase de l'exemple, apareix en l'argument 从资结构 (*en l'estructura d'inversió*) amb un rol AM-DIS del verb 看 (*veure*) i no representa un esdeveniment, ja que, juntament amb *estructura*, representa una entitat des de la qual alguna cosa pot ser vista. Els rols semàntics ajuden a distingir quines ocurrences de *inversió* són esdeveniments.

(130) TempEval-2, file: chtb.0618 - sent.: 10

此外，日本是中国吸引外资、引进技术的主要对象国。

Podem veure en l'estructura d'inversió que han crescut significativament els projectes amb alt contingut tecnològic.

Encara que va haver pocs casos i la cobertura no es va veure massa afectada, alguns esdeveniments que el *baseline* no reconeix van ser reconeguts correctament per TIRSemZH. L'exemple (131) mostra com els rols semàntics ajuden en el reconeixement dels esdeveniments *state*. En aquesta frase, 充满 (*téplena*) és un verb principal, que encapçala un predicat en l'oració i indica

un esdeveniment. Aquest no és reconegut pel *baseline* perquè en l'entrenament no apareix en aquest context de lema-POS o un de similar. No obstant això, utilitzant els rols semàntics, l'enfocament és conscient del fet que 充满 és un verb principal que regeix dos rols 社会公众 (*El públic*) A0, i 信心 (*confiança*) A1. A l'entrenament, molts esdeveniments es troben en aquesta situació semàntica i, per tant, la probabilitat del model per a reconèixer-los augmenta.

- (131) TempEval-2, file: chtb.0628 - sent.: 8
社会公众对宏观调控和经济发展充满信心。
El públic té plena confiança en el macro-control i el desenvolupament econòmic

Molts dels errors de l'aproximació basada en rols semàntics (TIRSemZH) van ser deguts a errors de l'eina SRL (la F1 de l'eina SRL és al voltant del 77%). Per exemple, el verb 减低 (*reduir*), que representa un esdeveniment en una frase del test, no és considerat un verb principal per l'eina SRL, i l'esdeveniment es perd. En el cas contrari, l'eina d'SRL etiqueta 贸易 (*comerciar*) com a verb principal en frases on representa un substantiu no eventiu i l'aproximació produeix un fals positiu.

En aquesta secció, hem analitzat la contribució de les característiques semàntiques de TIPSem en una avaluació multilingüe per a l'espanyol, italià i xinès. S'ha demostrat que aquestes característiques són útils per al processament de la informació temporal en diferents idiomes.

4.3 Conclusions

Per demostrar la nostra hipòtesi, hem mesurat el rendiment de TIPSem, i l'efecte de l'aplicació dels trets semàntics proposats per al processament d'informació temporal ha sigut analitzat en diferents idiomes: anglès, espanyol, italià i xinès.

TIPSem ha sigut avaluat amb els conjunts de dades públics TimeML utilitzats en l'exercici d'avaluació internacional TempEval-2. A més, els resultats han sigut comparats amb els obtinguts per un *baseline* no semàntic, TIPSem-B, que només utilitza característiques morfosintàctiques. La taula 4.32 resumeix els resultats obtinguts per a l'anglès¹⁵.

¹⁵Vegeu els resultats detallats d'una avaluació creuada en l'apèndix C.

| Element | Task | F1 | MRER_S (10-fold) | SSC |
|---------|-------------------------------|------|------------------|-------|
| TIMEX | recognition F1 | 0,88 | 26% | 99.5% |
| | classification <i>PA</i> | 0,98 | 19% | 97.5% |
| | normalization <i>PA</i> | 0,84 | 19% | 99.5% |
| EVENT | recognition F1 | 0,85 | 11% | 99.5% |
| | classification <i>PA</i> | 0,79 | 0% | - |
| TLINK | event- <i>timex</i> <i>CA</i> | 0,68 | 5% | 90% |
| | event-DCT <i>CA</i> | 0,82 | 3% | 90% |
| | main-events <i>CA</i> | 0,61 | 3% | 99.5% |
| | sub-events <i>CA</i> | 0,66 | 2% | 95% |

Taula 4.32: Resum dels resultats de TIPSem per a l'anglès (MRER_S: reducció mitjana de l'error relatiu introduït per la semàntica, SSC: Grau de confiança en la millora estadísticament significativa)

Dels resultats obtinguts, concloem que:

- *Les característiques semàntiques proposades són apropiades per a abordar el problema.* Per a l'anglès, l'aplicació de la semàntica introdueix una millora positiva i estadísticament significativa sobre el *baseline*. A més, els diferents tipus de semàntica aplicats ofereixen avantatges complementaris. Per tant, la seua combinació és útil i obté el millor rendiment. L'aplicació de la semàntica millora significativament el reconeixement de *timex* (26% de reducció relativa mitjana de l'error, 0,88 F1) i esdeveniments (11% de reducció relativa mitjana de l'error, 0,85 F1). En particular, la semàntica ajuda en el reconeixement de *timex* ambigües a nivell morfosintàctic i dels esdeveniments nominals. Els trets semàntics obtinguts a partir de TIMEK beneficien la classificació de *timex* (19% de reducció relativa mitjana de l'error, 0,98 F1) i la normalització (19% reducció relativa mitjana d'errors, 0,84 F1). L'ús de la semàntica temporal, la semàntica lèxica i els rols semàntics augmenta el rendiment en la categorització de relacions temporals: event-*timex* (5% de reducció relativa mitjana de l'error, 0,68 F1), event-DCT (3% de reducció relativa mitjana de l'error, 0,82 F1), esdeveniments principals (3% de reducció relativa mitjana de l'error, 0,61 F1), i esdeveniments subordinats (2% de reducció relativa mitjana de l'error, 0,66 F1). La classificació d'esdeveniments és l'única tasca que no ha sigut millorada pels trets semàntics (0% de reducció relativa mitjana de l'error, 0,79 F1), però s'han proposat mètodes alternatius per a aprofitar la semàntica en aquesta tasca.
- *S'obté un alt rendiment.* En comparació amb l'estat de la qüestió (SOA) per a l'anglès, TIPSem obté un resultat molt competitiu en

totes les subtasques, la qual cosa assenyala que el nostre enfocament semàntic ofereix un alt rendiment. En particular, TIPSem supera el rendiment del SOA en la tasca de processament d'esdeveniments (+0,05 F1), la qual cosa és una de les principals aportacions d'aquesta tesi (0,85 F1 en el processament d'esdeveniments).

- *La nostra aproximació es útil per a diferents llengües.* Les conclusions descrites han sigut confirmades en tots els idiomes inclosos en l'avaluació multilingüe: espanyol, italià i xinès. TIPSem supera el *baseline* i el rendiment del SOA en aquests idiomes. A més, fins on nosaltres sabem, TIPSem és la primera aproximació al processament d'esdeveniments per a l'espanyol i el xinès.
- *Cal més coneixement.* Pel que fa als errors que TIPSem deixa sense resoldre, a més dels procedents de les eines de PLN utilitzades o errors humans d' anotació, aquests són casos que requereixen una anàlisi lingüístic a un nivell superior a la semàntica. Per exemple, alguns esdeveniments nominals i *timex* requereixen pragmàtica, i algunes *timex* requereixen la resolució de l'anàfora temporal. A més, les relacions temporals sovint requereixen un raonament causal, l'anàlisi del discurs i el coneixement del món.

El capítol següent presenta una avaluació extrínseca de TIPSem a través de la seua aplicació en una interfície gràfica basada en el temps.

Capítol 5

Aplicació: Una interfície gràfica a la informació temporal - Time-Surfer

En el capítol anterior hem presentat una avaluació intrínseca del rendiment del nostre enfocament de processament de la informació temporal (TIPSem). Ara, passem a una avaluació extrínseca de TIPSem a través de la seua aplicació a una tasca diferent. En particular, TIPSem s'ha aplicat per a crear una interfície gràfica dinàmica a la informació temporal.

La correcció de la informació temporal extreta dels textos a través de mitjans manuals o computacionals pot ser analitzada extrínsecament amb les seues diferents aplicacions.

En el PLN, el processament de la informació temporal s'ha aplicat amb èxit a la recuperació d'informació (Alonso et al., 2007a), a la cerca de respostes (Saquete et al., 2009), i, més recentment, a la construcció d'interfícies gràfiques per a accedir a la informació (Alonso et al., 2010). La Informació temporal millora les capacitats dels sistemes quan aquests tracten amb variables temporals en consultes, qüestions temporals complexes o representacions gràfiques del temps.

TIPSem s'ha aplicat per a crear una interfície gràfica innovadora per a explorar el contingut dels documents basada en la dimensió temporal. Hem anomenat a aquesta interfície **Time-Surfer** (Llorens, Saquete, Navarro, & Gaizauskas, 2011b). Per a representar els aspectes temporals d'un document, Time-Surfer processa el document utilitzant TIPSem. Això relaciona l'èxit de Time-Surfer amb el rendiment de TIPSem.

Time-Surfer pot ser vist com una millora d'un sistema de recuperació

d'informació (IR) d'ús específic.

La quantitat d'informació digitalitzada ha crescut fins a sobrepassar els límits humans. L'IR estudia mètodes per recuperar els documents més rellevants en una col·lecció donada una consulta de cerca de l'usuari. No obstant això, aquests documents són sovint grans i contenen molta més informació que no interessa a l'usuari. Per tant, s'hi han inclòs mètodes per a ajudar l'usuari a localitzar la informació que desitja en els documents més rellevants. Els motors de cerca actuals mostren els fragments més rellevants de cada resultat de la cerca, però recentment el valor de la dimensió temporal en l'IR s'ha posat també en relleu (Alonso et al., 2007a).

La informació temporal està present no només en les metadades, com a data de creació o modificació, sinó també en el contingut del document –en les expressions temporals i esdeveniments. L'objectiu del Time-Surfer és presentar una interfície gràfica per a accedir a la informació dins d'un document, utilitzant tècniques avançades de representació temporal i de navegació. En concret, el Time-Surfer es basa en la identificació de les *timex*, els esdeveniments i les relacions temporals i després utilitza aquesta informació per a proporcionar als usuaris un accés dinàmic als textos basat en el text.

Aquest capítol està estructurat de la següent manera. En primer lloc, es revisen les principals representacions gràfiques del temps i els treballs relacionats sobre interfícies gràfiques temporals. A continuació, es descriuen l'arquitectura i les característiques de Time-Surfer. Finalment, es presenta l'avaluació del Time-Surfer i es discuteixen els resultats.

5.1 Representant el temps gràficament

Hi ha dues formes principals en què s'ha representat gràficament la informació temporal: **línies de temps** i **grafs**.

D'una banda, la línia de temps és la representació més intuïtiva, perquè això implica l'associació d'esdeveniments a referències calendàriques en una línia homogènia. La línia representa el temps absolut Newtonià –el que es mostra en els rellotges i calendaris. El problema sorgeix quan la informació temporal és incompleta, difusa o no específica. En aquest cas els esdeveniments no poden ser adequadament col·locats sobre la línia de temps.

D'altra banda, els grafs de temps resolen el problema de les línies de temps per a mostrar informació temporal incompleta, difusa o no específica. Els gràfics poden representar l'ordre relatiu d'esdeveniments, si hi ha un temps d'ancoratge. El problema és que, encara que els grafs també poden

5.1. REPRESENTANT EL TEMPS GRÀFICAMENT

representar els ancoratges temporals d'esdeveniments a les dates del calendari, perden intuïció i claredat. Per a entendre'n els avantatges i inconvenients, considerem l'exemple (132)¹.

(132) El pilot s'ejectà abans de l'accident. *Nota: l'accident passa al juny del 2000*

Com es mostra en la figura 5.1, si tractem de representar els fets en una línia de temps podem situar *accident* al juny del 2000 (6/00), però no hi ha un lloc específic per a posar el esdeveniment *ejectà*, perquè l'única pista que tenim és que és anterior a *accident*. Setzer i Gaizauskas argumentaren que la col·locació de *ejectar* en algun lloc de la línia de temps, inevitablement, fa que s'associe amb la data de baix, mentre que només s'hauria de considerar en relació amb *accident*.

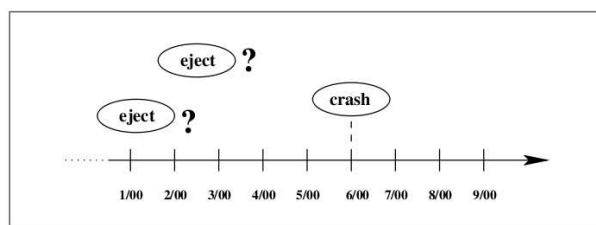


Figura 5.1: Problema de la representació de línia de temps I

A més, com es mostra en la figura 5.2, si tenim un grup d'esdeveniments (és a dir, B, C, D) que poden ser col·locats en relació amb un altre esdeveniment (és a dir, A), però no sabem la relació temporal entre si, hem de situar B, C, D en la línia de temps. Això dóna la impressió que sabem la seua relació i el moment en el qual ocorren.

El problema és que les línies temporals forcen a detallar el moment en que ocorren els esdeveniments i això sovint no es detalla en el text. Tanmateix, això no és un problema per als grafs de temps, com es mostra en la figura 5.3.

A més del que Setzer i Gaizauskas argumenten, veiem un altre problema de la representació de línia de temps. Suposant que, per exemple, la línia de temps tinga una granularitat de minuts, si *accident* passa al juny del 2000, hem de decidir on posar l'*accident* en una precisió de minuts (**problema de granularitat imprecisa**). Podríem assumir una posició per defecte,

¹Exemple tret de Setzer i Gaizauskas (2001).

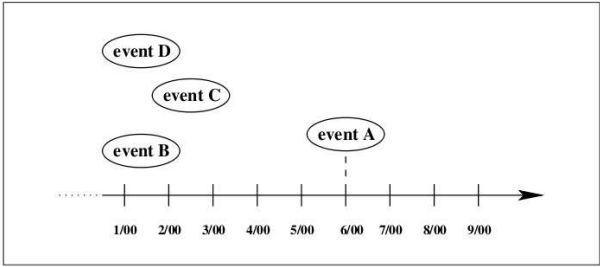


Figura 5.2: Problema de la representació de línia de temps II

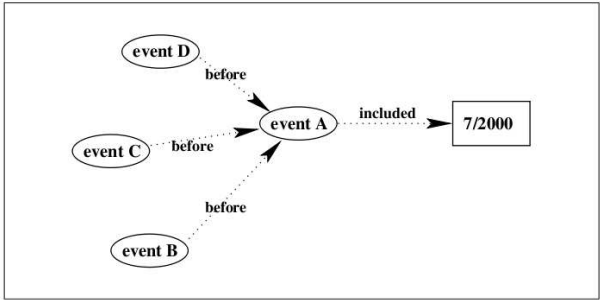


Figura 5.3: Representació amb graf temporal

com ara 2000-01-06 00:00. Però fa l'efecte que sabem l'ancoratge exacte de l'esdeveniment en un punt precís del calendari (en minuts).

Aquests problemes marquen una preferència en favor dels grafs de temps per a les representacions temporals formals. No obstant això, encara que els grafs són la millor opció per a la representació formal, des del nostre punt de vista, quan l'usuari final de la representació és un ésser humà, les línies temporals han de ser reconsiderades com a representació bàsica a causa de la seua intuïció. Creiem que l'aplicació final de la representació dicta l'elecció. Els grafs proporcionen una representació exacta, però complicada, i les línies temporals proporcionen una representació inexacta, però intuïtiva.

5.2 Treballs relacionats

Treballs recents de recerca en l'IR demostren la importància de la dimensió temporal (Alonso et al., 2007a; Berberich, Bedathur, Alonso, & Weikum, 2010). La majoria d'aquestes obres utilitzen línies de temps com la representació temporal bàsica (Alonso, Gertz, & Baeza-Yates, 2007b; Alonso et al., 2010). Senyals clares d'aquesta preferència per la línia de temps en aplicacions d'usuari són Google Timeline², i Yahoo's Time Explorer³ basat en SIMILE (www.simile-widgets.org).

Els enfocaments esmentats representen els resultats de cerca en una línia de temps utilitzant les expressions temporals que figuren en els documents o dins de les metadades (per exemple, la data de publicació). Només Alonso et al. (2010) representa els principals esdeveniments en el contingut dels documents recuperats, però els extreuen manualment usuaris humans.

El nostre enfocament es diferencia d'aquestes propostes en els aspectes següents. En primer lloc, Time-Surfer se centra a representar amb precisió tots els esdeveniments continguts en documents de grans dimensions i ofereix funcions de recerca, comparació i de navegació amb zoom dinàmic en el temps. En segon lloc, Time-Surfer utilitza eines computacionals per a extreure les referències a esdeveniments, expressions temporals i relacions temporals, de manera que tot el procés és automàtic.

5.3 Time-Surfer: arquitectura i característiques

L'arquitectura del Time-Surfer consta de tres passos principals, que es descriuen a continuació i es resumeixen gràficament en la figura 5.4.

²<http://newstimeline.googlelabs.com/>

³<http://fbmya01.barcelonamedia.org:8080/future/>

El primer pas consisteix en l'extracció de les expressions temporals, esdeveniments i relacions temporals d'un document d'entrada. Els recents avanços en el camp de l'extracció d'informació (IE) temporal s'han traduït en la proliferació de sistemes capaços d'extraure aquests elements seguint esquemes d'anotació com TimeML (Pustejovsky et al., 2003). En particular, en aquest pas emprem la contribució central d'aquesta tesi (TIPSem). Time-Surfer utilitza el sistema de TIPSem (Llorens et al., 2010b) per a anotar en els documents d'entrada amb TimeML.

En el segon pas s'extrauen els grups d'esdeveniments (TimeML) que estan vinculats a la mateixa referència de temps en grups d'esdeveniments (*event groups*). La classe d'esdeveniment, la ubicació en el text original, els participants i el context textual es guarden en un format estructurat per a ser usats posteriorment. El format estructurat es guarda en el format JSON (JavaScript Object Notation)

Finalment, en el tercer pas, aquesta informació es carrega en la nostra interfície gràfica interactiva. Per a desenvolupar aquesta interfície basada en la web, es va adaptar i ampliar la biblioteca jQuery Flot⁴ amb noves funcions de representació, navegació i cerca. Les tecnologies web que s'utilitzen són PHP, JavaScript i HTML.

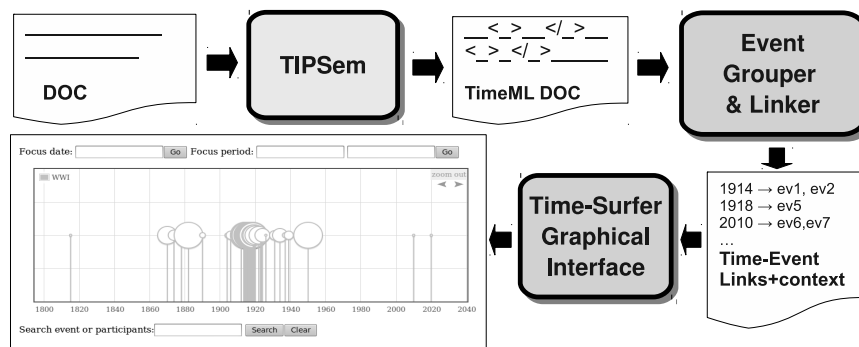


Figura 5.4: Arquitectura Time-Surfer

La interfície ofereix les capacitats interactives següents:

- **Visió general - Línia del temps:** Cada grup d'esdeveniments està representat per una bombolla que té una posició que depèn del moment en què tenen lloc els esdeveniments, i el radi depèn de la quantitat

⁴<http://flot.googlecode.com>

d'esdeveniments que conté. Això dona una visió general de la distribució temporal dels esdeveniments del document, i il·lustra els límits de temps i períodes importants. La figura 5.4 mostra aquesta vista de l'article de la Viquipèdia de la Primera Guerra Mundial.

- **Explorar (*surfing*):** La interfície permet a l'usuari navegar per la línia de temps.
 - **Assenyalar amb el punter:** En passar per damunt d'un grup d'esdeveniments, es mostra informació detallada (data, nombre d'esdeveniments, llista dels esdeveniments en el seu context).
 - **Clicar:** En fer clic en un grup d'esdeveniments, la informació anterior es presenta en una capa nova. A partir d'aquí, l'usuari pot fer clic en una frase específica i anar al text del document original.
 - **Arrossegar:** L'usuari pot desplaçar-se cap arrere i cap davant al llarg de la línia de temps de forma intuïtiva arrossegant-la i deixant-la anar (drag&drop).
 - **Fer zoom:** Utilitzant la roda del ratolí l'usuari pot fer zoom en el temps. Aquest zoom dinàmic permet l'exploració de grups d'esdeveniments molt propers en el temps. Aquesta és una de les característiques que diferencia el nostre enfocament dels treballs relacionats.
- **Cercar:** Time-Surfer també inclou formularis per a buscar una data (per exemple, 1916) o període (per exemple, 1914-1918), i la recerca d'esdeveniments (per exemple, *batalla*), participants de l'esdeveniments (per exemple, *Alemanya*) o les relacions entre participants (per exemple, *Hitler - Mussolini*). Si s'introdueix una consulta, les bombolles que contenen els fets rellevants es pinten, mentre que les altres es fan semitransparents. En el text, els casos trobats es destaquen, així com les frases que els continguin (vegeu la figura 5.5).
- **Comparació de documents en el temps:** La interfície permet comparar diversos documents en el temps, mantenint les característiques dinàmiques anteriors. Ací els grups d'esdeveniments de cada document es mostren amb un color diferent, però es col·loquen en la mateixa línia de temps (vegeu la figura 5.6).

Time-Surfer s'ha aplicat a un conjunt d'articles d'història i biogràfics de Viquipèdia per tal de demostrar les característiques descrites –vegeu la demo en línia en <http://gplsi.dlsi.ua.es/demos/TIMEE/Time-Surfer/>.

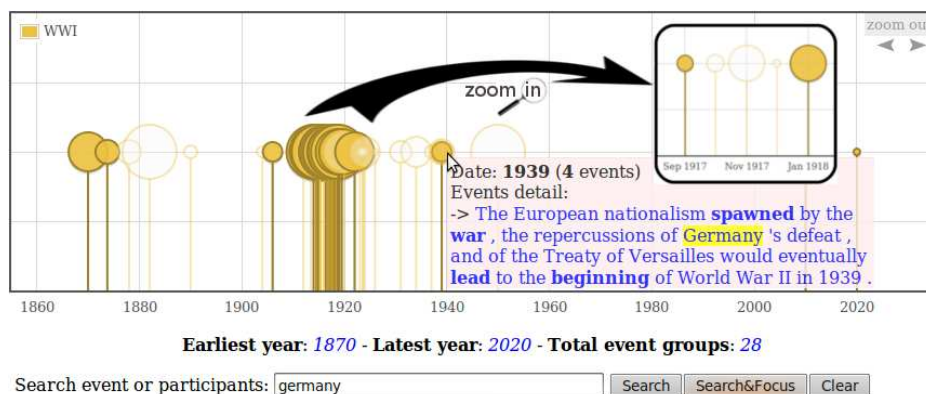
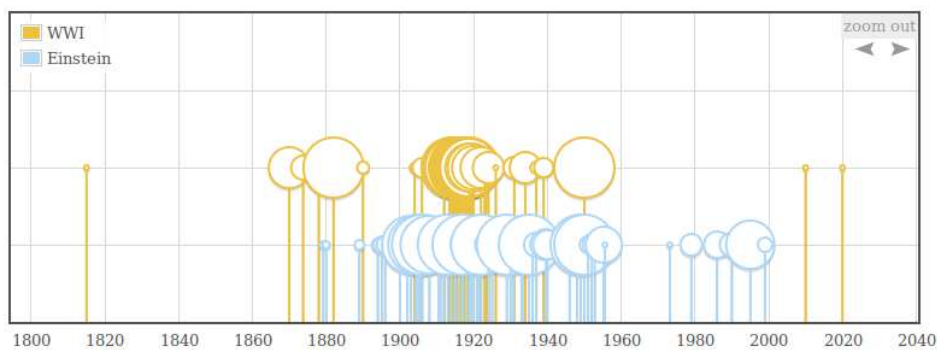
Figura 5.5: Buscant *Alemanya* en WWI, assenyalant 1939, i fent zoom

Figura 5.6: Comparació de WWI i Einstein

5.4 Avaluació de Time-Surfer

Per tal de comprovar la utilitat i els avantatges de la interfície Time-Surfer, es proposa una avaluació orientada a l'usuari. L'avaluació es basarà en el rendiment dels usuaris sobre l'ús del Time-Surfer en contrast amb altres interfícies, i en el seu grau de satisfacció. L'objectiu d'aquesta avaluació és doble: (i) mesurar els avantatges que el Time-Surfer mostra en comparació amb les interfícies web tradicionals (Viquipèdia), i (ii) mesurar la satisfacció dels usuaris amb el Time-Surfer. A més, l'avaluació inclou els idiomes castellà i anglès; i serveix com a avaluació extrínseca de TIPSem.

L'experiment que proposem es pot resumir en dues parts principals:

- **Avantatges de Time-Surfer:** L'avaluació consisteix en un conjunt de 10 preguntes temporals sobre algun article de Viquipèdia que han de ser respostes per la persona avaluada –5 utilitzant la interfície de Time-Surfer i 5 per mitjà de la interfície de Viquipèdia–. En primer lloc, s'explica durant 10 minuts com s'utilitza el Time-Surfer a la persona que serà avaluada. Les primeres cinc preguntes es mostren i es mesura el temps que empra l'usuari per a donar la resposta correcta usant Viquipèdia. Després, es demana al mateix individu que utilitzi el Time-Surfer per a respondre a les cinc preguntes següents i es mesura el temps també. Per a evitar la dependència dels resultats amb les competències individuals, es van avaluar grups d'individus. A més d'evitar la dependència dels resultats sobre la distribució de les preguntes, aquestes s'intercanvien en cada avaluació –les preguntes que es responien amb Time-Surfer es canvien a Viquipèdia, i viceversa.
- **Test de satisfacció:** Després d'acabar la part anterior, als individus avaluats se'ls demana omplir un test de satisfacció que es mostra en (133). La prova consta de preguntes amb quatre opcions de resposta ordenades en ordre ascendent de satisfacció (1 molt insatisfet, 2 insatisfet, 3 satisfet, i 4 molt satisfet). Una selecció de quatre evita la selecció d'un grau de satisfacció no polaritzat.

(133) Time-Surfer: Test de Satisfacció

Per favor, seleccioneu el vostre nivell de satisfacció d'acord amb les preguntes següents.

1 .- Creieu que l'aplicació Time-Surfer és útil com a representació temporal de texts històrics i biogràfics?

☐ inútil ☐ no molt útil ☐ un poc útil ☐ molt útil

2 .- Creieu que el Time-Surfer és útil com a interfície per a explorar i accedir a aquesta informació?

☐ inútil ☐ no molt útil ☐ un poc útil ☐ molt útil

3 .- Us és fàcil trobar la informació en el temps mitjançant Time-Surfer?

☐ inútil ☐ no molt útil ☐ un poc útil ☐ molt útil

4 .- Creieu que el Time-Surfer és útil per a obtenir una idea de la distribució de la informació en el temps?

☐ inútil ☐ no molt útil ☐ un poc útil ☐ molt útil

5 .- Usaríeu Time-Surfer a la feina per a buscar o explorar la informació temporal?

__ mai __ poques vegades __ de vegades __ moltes vegades

L'experiment descrit s'ha aplicat a estudiants de doctorat de diferents àrees. Els estudiants de doctorat representen un grup de persones amb necessitats d'informació altes i que sovint utilitzen les tecnologies de la informació per a buscar i explorar. Per a l'avaluació, es van seleccionar els estudiants de doctorat de Ciències de la Computació i de Filologia Espanyola. D'una banda, els científics de la computació tenen bones habilitats d'adaptació a les noves interfícies humà-màquina que mai han usat abans, com Time-Surfer. D'altra banda, els experts de la filologia espanyola tenen bones habilitats en la cerca en textos electrònics llargs, com alguns articles de la Viquipèdia, i menor capacitat d'adaptació a les interfícies noves com Time-Surfer. Els resultats d'aquesta avaluació ens donaran les claus per a millorar la versió actual de la interfície i fer-la més útil i agradable per als usuaris potencials.

Avaluació en els estudiants de doctorat de ciències de la computació

Un total de sis estudiants de doctorat de Ciències de la Computació van ser avaluats seguint la prova descrita. En particular, les deu (5+5) preguntes plantejades en aquest experiment es presenten en (134).

(134) Time-Surfer: Preguntes per als Científics de la Computació

(La Guerra Civil Espanyola Viquipèdia l'article)

1 .- Quin és el primer esdeveniment en el temps que apareix en l'article?

2 .- Què va passar en 1934-10-01?

3 .- Quines dates estan relacionades amb Hitler?

4 .- Quines són les dates en què Hitler i Mussolini estan relacionats?

(J.F.Kennedy-Einstein articles de Viquipèdia)

5 .- Què va passar amb el fill de Kennedy, quan Albert Einstein va ser nomenat *físic més gran de tots els temps*?

(Primera Guerra Mundial, l'article de Viquipèdia)

6 .- Quin és l'últim esdeveniment en el temps que apareix en l'article?

7 .- Què va passar en 1916-08-27?

8 .- Quines dates estan relacionades amb Itàlia?

9 .- Quines dates estan relacionades amb Itàlia i Alemanya?

(J.F.Kennedy-Einstein articles de Viquipèdia)

10 .- Què li va passar a Kennedy després que Einstein guanyara la medalla de Franklin?

Com es pot observar en el conjunt de preguntes, les preguntes 1 a 5 són anàlogues a les preguntes 6 a 10. Aquestes representen diferents tipus de preguntes temporals:

- **Preguntes d'ordre temporal:** Les preguntes 1 i 6 requereixen que l'usuari trobe les primeres o les últimes referències temporals en un document de text i obtenir-hi l'esdeveniment associat. La dificultat d'aquestes qüestions recau en el fet que els documents sovint no estan ordenats en una línia de temps lineal, és a dir, la primera frase no sempre conté el primer esdeveniment en el temps del document ni l'última frase sempre conté els últims. En el text en llenguatge natural pot haver-hi *flash-backs* i referències futures puntuals. Quan s'utilitza Viquipèdia (informació textual) l'única manera que un usuari té per a trobar de manera fiable els primers o els últims esdeveniments és llegir tot el document i retenir la primera o l'última referència temporal (data) i els seus esdeveniments associats. Per contra, quan s'utilitza Time-Surfer, aquesta tasca ja està feta i els esdeveniments es mostren ordenats en una línia de temps gràfica.
- **Preguntes de dates:** Les preguntes 2 i 7 requereixen que l'usuari trobe l'esdeveniment associat a una data explícita. La dificultat d'aquestes qüestions recau en el fet que els documents poden contenir una data en una varietat de formats explícits i implícits. Per exemple, tenint en compte que avui és 24 de febrer de 2001, la data 12/24/2000 es pot expressar com fa dos mesos, el 24 de desembre de 2000, etc. Quan s'utilitza la Viquipèdia, l'usuari ha d'endevinar quin és el format de la data i trobar-la amb els mitjans de recerca de text. Per contra, quan s'utilitza Time-Surfer les dates es poden buscar directament, independentment del format que tinguen en el text.
- **Preguntes d'esdeveniment o participant (I):** Les preguntes 3 i 8 requereixen que l'usuari trobe les dates associades a un esdeveniment o un participant de l'esdeveniment (per exemple, *Hitler*, *Itàlia*). En la recerca de text normal, l'usuari ha de buscar en diverses ocasions i llegir les oracions per a trobar les possibles dates vinculades. No obstant això, amb TIPSem, una vegada que la consulta s'insereix, les dates es destaquen en la línia de temps.
- **Preguntes d'esdeveniment o participant (II):** Les preguntes 4 i 9 requereixen que l'usuari trobe les dates associades a dos esdeveniments o participants que apareixen junts (per exemple, *Hitler i Mussolini*,

Itàlia i Alemanya). TIPSem ofereix formes específiques de cerca per a trobar relacions, mentre que la cerca de text normal no ho fa.

- **Preguntes combinades:** Les preguntes 5 i 10 requereixen que l'usuari combine informació de dues fonts (documents). Aquest tipus de preguntes tenen per objecte avaluar la utilitat de les funcions de comparació de documents de TIPSem amb els navegadors de text tradicionals.

La figura 5.7 mostra el temps mitjà en segons (temps aproximat) que els individus tarden a donar la resposta correcta per a cada pregunta amb Time-Surfer i Viquipèdia. Els dos gràfics es diferencien perquè els individus que utilitzen la Viquipèdia en les preguntes 1-5 utilitzen Time-Surfer en les preguntes 6 a 10 i viceversa.

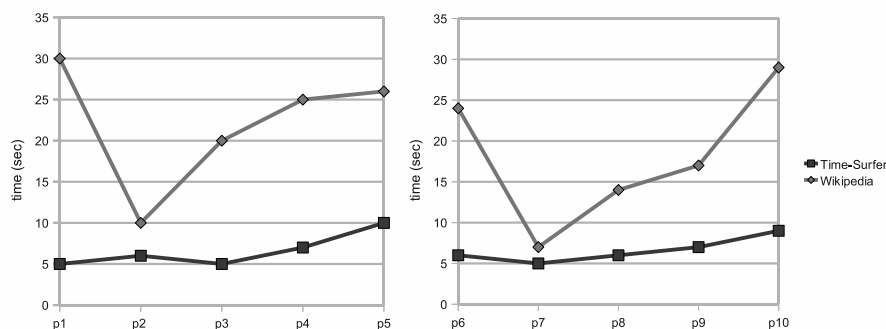


Figura 5.7: Time-Surfer *vs.* Viquipèdia (mitjana de científics de la computació)

Com es pot observar, de mitjana, els usuaris van respondre més ràpid quan utilitzaren Time-Surfer, com s'esperava. La diferència ha sigut més accentuada en les preguntes d'ordre temporal i en les combinades. En les preguntes de dates la diferència és menor. Es van analitzar aquests casos, i s'ha advertit que les dates en Viquipèdia solen estar en format explícit i comú, de manera que les persones les poden trobar ràpidament. Probablement, si les dates foren implícites, la diferència en el temps de resposta, en comparació amb el Time-Surfer, seria més gran.

La figura 5.8 informa sobre el grau de satisfacció mitjà dels individus avaluats després de la prova de Time-Surfer. Recordem que la prova de satisfacció (vegeu l'exemple 133) permet quatre nivells de satisfacció (1 molt

insatisfet, 2 insatisfet, 3 satisfet, 4 molt satisfet), sent la mitja (2,5) el valor no polaritzat.

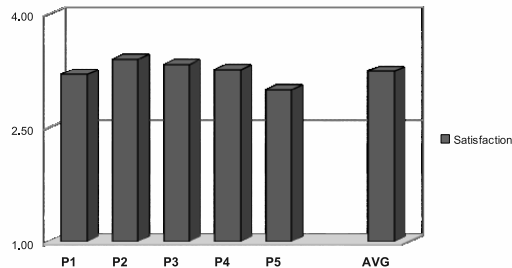


Figura 5.8: Time-Surfer satisfacció (científics de la computació)

Les puntuacions de satisfacció superen la mitjana (2,5) per a totes les preguntes; per tant, el grau de satisfacció obtingut és positiu. La satisfacció per pregunta mostra que la major satisfacció dels usuaris és la utilitat de Time-Surfer com una interfície per a explorar la informació en el temps, mentre que la menor es troba en la freqüència en què l'emprarien en les seues tasques de recerca diària. A més, els usuaris estan satisfets amb la representació temporal, amb com és d'amigable per a l'usuari, i amb la vista per defecte que Time-Surfer presenta d'un document.

Avaluació en estudiants de doctorat de filologia hispànica

Un total de quinze estudiants de doctorat de Filologia Espanyola van ser avaluats amb la prova descrita. En particular, les deu (5+5) preguntes plantejades en aquest experiment es presenten en (135). Vam adaptar les preguntes al camp de la literatura espanyola i les vam traduir a l'espanyol perquè foren més familiars als individus avaluats. Se'ls van fer preguntes temporals sobre les biografies d'escriptors espanyols de la Viquipèdia (Cervantes, Lope de Vega, Garcilaso, Vargas Llosa, etc.) Encara que el tema canvia, el tipus de preguntes utilitzat és el mateix per a l'avaluació dels estudiants de doctorat de Ciències de la Computació (primer i últim esdeveniment, esdeveniment d'una data específica, els esdeveniments relacionats amb una consulta de múltiples documents, etc.).

(135) Time-Surfer: Preguntes per als filòlegs hispànics

- 1 .- Quin és l'esdeveniment més recent en el temps que apareix sobre Lope de Vega?
- 2 .- Què li va passar a Garcilaso entre 1528 i 1529?

CAPÍTOL 5. APLICACIÓ: TIME-SURFER

- 3 .- Què passa amb Cervantes el 1584-12-12?
- 4 .- On vivia Márquez quan Vargas Llosa va escriure *Elogio de la madrastra*?
- 5 .- Què va passar amb Márquez quan Vargas Llosa va nèixer?
-
- 6 .- Quin és l'últim esdeveniment en el temps que apareix sobre Cervantes?
- 7 .- Quina és la data dels primers esdeveniments en el temps que apareix en la biografia de Quevedo?
- 8 .- Què li va passar a García Lorca el 1933-02-11?
- 9 .- Quines novel·les va escriure Galdós en els anys anteriors al naixement de Lorca?
- 10 .- Què li va passar a Cervantes quan Quevedo va nèixer?

Una vegada més, les preguntes responen als tipus descrits abans (ordre temporal, data, etc.)

La figura 5.9 mostra el temps mitjà en segons (temps aproximat) que els individus tarden amb Time-Surfer i Viquipèdia.

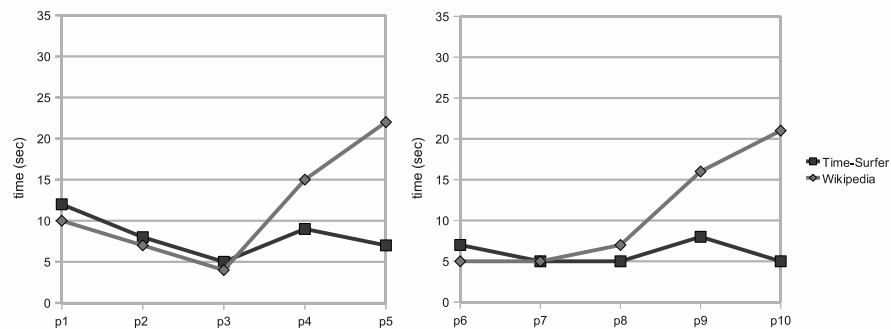


Figura 5.9: Time-Surfer *vs.* Viquipèdia (Filologia Hispànica)

Com es pot observar, de mitja, els usuaris van respondre més ràpid utilitzant Time-Surfer, com s'esperava. La diferència ha sigut més accentuada en les preguntes combinades. Una vegada més, les preguntes de data obtenen una diferència de temps més baixa a causa de la forma explícita de les dates en el text. Una diferència interessant entre els resultats dels científics de la computació i els estudiants de Filologia es troba en les qüestions d'ordre temporal. Els estudiants de Filologia les van respondre ràpid amb Viquipèdia. Mentre que en Time-Surfer els usuaris encara estaven acostumant-se a la interfície (era la primera pregunta que havien de respondre amb la interfície),

els usuaris de Viquipèdia van endevinar la resposta utilitzant els seus coneixements previs sobre la biografia dels escriptors suposant que coincidiria, per exemple, amb les dates de naixement.

La figura 5.10 informa sobre el grau de satisfacció mitjà dels individus avaluats després de la prova de Time-Surfer.

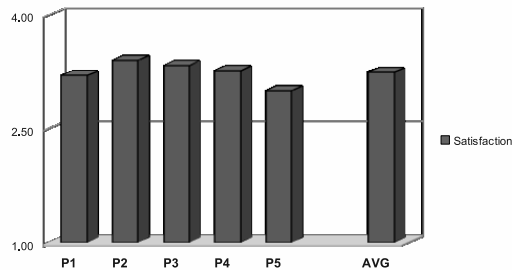


Figura 5.10: Satisfacció sobre l'ús de Time-Surfer (Filologia Hispànica)

La mitjana de les puntuacions de satisfacció és molt similar a la reportada per a científics de la computació. Totes les preguntes mostren un grau de satisfacció positiu (> 2.5). Els graus de satisfacció per pregunta confirmen que la major satisfacció dels usuaris és la utilitat de Time-Surfer com a interfície per a explorar la informació en el temps, mentre que la menor és la freqüència en què l'emprarien en les seues tasques de recerca diària.

5.5 Conclusions

Per tal d'avaluar TIPSem extrínsecament, s'ha aplicat la creació d'una interfície gràfica al contingut dels documents basada en la dimensió temporal (Time-Surfer).

L'objectiu del Time-Surfer és permetre als usuaris cercar i explorar la informació relativa a un període específic, esdeveniment o participant d'esdeveniment en un document. A diferència de les actuals interfícies gràfiques, Time-Surfer no només es basa en la detecció automàtica de *timex*, sinó també dels esdeveniments i les relacions temporals. A través d'una interfície de línia de temps amb zoom, ofereix als usuaris una imatge dinàmica de la distribució temporal dels esdeveniments d'un document.

Time-Surfer s'ha analitzat mitjançant una avaluació orientada a l'usuari usant articles històrics i biogràfics de Viquipèdia. S'ha mesurat el temps que els usuaris empenen per a respondre preguntes i el grau de satisfacció amb el

Time-Surfer en comparació amb interfícies d'accés a la informació clàssiques (com la Viquipèdia).

Els resultats de l'avaluació mostren que amb un temps d'adaptació curt, l'usuari es beneficia de la utilització del Time-Surfer per a la recerca i exploració de la informació temporal. A més, el grau de satisfacció obtingut mostra que el Time-Surfer és útil, fàcil d'usar i aplicable per als individus avaluats.

Tenint en compte que cada error de TIPSem en el processament de la informació temporal es propaga a Time-Surfer, els resultats obtinguts demostren que el rendiment de TIPSem sobre text general (històric i biogràfic), és prou alt per a servir d'entrada per a aquesta aplicació.

Capítol 6

Conclusions

Aquest capítol descriu les contribucions d'aquesta tesi i les orientacions futures d'aquest treball de recerca. A més, aquest capítol inclou un resum de la producció científica (publicacions) i la producció de programari (aplicacions desenvolupades) elaborades durant el desenvolupament d'aquesta tesi.

6.1 Contribució

Aquesta tesi es centra en el processament de la informació temporal, que s'emmarca en l'àrea de processament del llenguatge natural (PLN). El processament de la informació temporal implica la interpretació automàtica i el raonament sobre els aspectes temporals de la llengua, cosa que és crucial per a altres aplicacions del PLN, com ara el resum automàtic, la recuperació de la informació i la cerca de respostes.

La investigació sobre el processament de la informació temporal ha evolucionat a partir de diferents estratègies racionalistes formals a l'estratègia empírica basada en corpus, que consisteix en l'anotació de corpus d'acord amb un esquema d'anotació temporal en particular. Actualment, l'esquema estàndard d'anotació temporal és TimeML a causa de la seua integritat i les millores que inclou respecte als seus predecessors. TimeML explicita les expressions lingüístiques que transmeten esdeveniments, temps i relacions temporals.

L'anotació automàtica en TimeML és una tasca complexa, que requereix coneixements lingüístics en el nivell semàntic, de discurs i pragmàtic. Els esdeveniments, les expressions temporals i les relacions temporals poden ser ambigües en el nivell morfosintàctic. No obstant això, la majoria dels plantejaments computacionals de l'estat de la qüestió per a l'anotació Ti-

meML automàtica només utilitzen informació morfosintàctica. Les poques propostes que inclouen semàntica es centren en l'ús de la semàntica lèxica, i només una d'aquestes inclou els rols semàntics (rols de FrameNet). A més, la influència de la semàntica en el rendiment dels sistemes no s'ha analitzat en profunditat.

Prenent aquestes afirmacions com a motivació, es presenta una aproximació semàntica per al processament de la informació temporal: el sistema TIPSem. La principal aportació de la nostra proposta als plantejaments de l'estat de la qüestió és l'aplicació de determinades característiques semàntiques. TIPSem és un enfocament híbrid (basat en aprenentatge automàtic i basat en regles), que inclou característiques semàntiques (semàntica lèxica, rols semàntics i semàntica temporal), a més de les característiques morfosintàctiques.

L'arquitectura de TIPSem es divideix en tres mòduls que afronten les principals subtasques del processament de la informació temporal: el processament d'expressions temporals (reconeixement, classificació i normalització), el processament d'esdeveniments (reconeixement i classificació) i el processament de relacions temporals (categorització).

L'enfocament es basa principalment en aprenentatge automàtic, però també inclou un recurs basat en regles (TIMEK) per a obtenir algunes característiques, així com per a normalitzar les *timex* a la norma ISO 8601. La part basada en aprenentatge automàtic inclou dues tècniques diferents i s'hi empen camps aleatoris condicionals (CRF) i màquines de suport vectorial (SVM).

La part més important del sistema és l'ús de la semàntica. Les característiques semàntiques definides per a cada tasca es resumeixen a continuació:

- *Processament de timex*: Les característiques de la lèxica semàntica són els quatre hiperònims més generals de cada paraula i algunes característiques específiques extretes de TIMEK (p. e.: *patró*). Les paraules que figuren en *timex* són sovint hipònims dels conceptes en anglès ***time***, ***time_period*** o ***time_unit***, que estan entre els quatre hiperònims més generals en la jerarquia. A més, moltes *timex* poden generalitzar-se en patrons (p. e.: *fa dos dies* = fa NUM TUNIT).

Pel que fa als rols semàntics, s'utilitzen dues característiques: *rol* i *rol+lema*. Les *timex* normalment estan contingudes en els rols temporals o, per a alguns verbs, en arguments numerats relacionats amb el temps. El rol en què figura una paraula i el seu lema ofereixen informació útil per a determinar si és probable que siga una *timex* o no

(p. e.: AM-TMP 1999, A1 1999, AM-TMP+reunió).

- *Processament d'esdeveniments.* Només els quatre primers hiperònims de WordNet s'utilitzen com a semàntica lèxica. Les paraules contingudes en *timex* són sovint hipònims dels conceptes en anglès *event* i *state*, que estan entre els quatre primers hiperònims en la jerarquia.

Pel que fa als rols semàntics, s'utilitzen quatre característiques: *rol*, *rol+paraula*, *rol+verb*, i *rol+top4hypers*. El rol semàntic d'una paraula i les seues dependències amb la paraula, el verb, i els quatre hiperònims més generals són importants per a reconèixer i classificar els esdeveniments que són ambigus en els nivells inferiors d'anàlisi del llenguatge (p. e.: en anglès *work* o *control*). Aquestes característiques haurien d'ajudar a generalitzar els contextos semàntics més i menys probables dels esdeveniments (p. e.: A1+dur_a_terme vs. A1+menjar).

- *Processament de relacions temporals.* Pel que fa a la semàntica lèxica, la classe d'esdeveniment s'utilitza per a categoritzar la relació temporal entre dos esdeveniments en una relació de subordinació. Pel que fa als rols semàntics, si un element relacionat es troba en una subordinació temporal, la conjunció s'inclou com una característica (p. e.: *quan, després, mentre*). Pel que fa a la semàntica temporal, s'utilitza *valdiff*. Aquesta característica captura la direcció temporal o la diferència entre la referència temporal de la relació i la referència temporal sintàcticament relacionada amb els esdeveniments de la relació (p. e.: [1999,1999] = *igual*, [1999,1996]=*posterior*) .

Aquestes característiques haurien de millorar la capacitat de generalització dels models i ajudar en les expressions que són ambigües en el nivell morfosintàctic.

TIPSem ha sigut ampliat de l'anglès a altres idiomes, a saber: espanyol, italià i xinès. En aquests idiomes, s'apliquen característiques semàntiques anàlogues. Un avantatge d'aquest enfocament és que, ja que està basat principalment en aprenentatge automàtic, només el preprocessament i el recurs TIMEK han de ser modificats.

Hem dut a terme una avaluació empírica per a provar la nostra hipòtesi. El rendiment de l'aproximació semàntica presentada s'ha mesurat, i l'efecte de l'aplicació de les característiques semàntiques proposades per al processament d'informació temporal s'ha analitzat en quatre idiomes diferents: anglès, espanyol, italià i xinès.

TIPSem ha sigut avaluat en els conjunts de dades TimeML públics utilitzats en l'exercici d'avaluació internacional TempEval-2. A més, els re-

sultats han sigut comparats amb els obtinguts per un baseline no semàntic (TIPSem-B) i els de l'estat de la qüestió.

Dels resultats obtinguts, concloem que:

- *Les característiques semàntiques proposades són apropiades per a abordar el problema.* Per a l'anglès, l'aplicació de la semàntica introdueix una millora positiva i estadísticament significativa sobre el baseline. A més, els diferents tipus de semàntica aplicats ofereixen avantatges complementaris. Per tant, la seua combinació és útil i obté el millor rendiment. L'aplicació de la semàntica millora significativament el reconeixement de *timex* (26% de reducció relativa mitjana de l'error, 0,88 F1) i esdeveniments (11% de reducció relativa mitjana de l'error, 0,85 F1). En particular, la semàntica ajuda en el reconeixement de *timex* ambigu en el nivell morfosintàctic i dels esdeveniments nominals. Els trets semàntics obtinguts a partir de TIMEK beneficien la classificació de *timex* (19% de reducció relativa mitjana de l'error, 0,98 F1) i la normalització (19% de reducció relativa mitjana d'errors, 0,84 F1). L'ús de la semàntica temporal, semàntica lèxica i rols semàntics augmenta el rendiment en la categorització de relacions temporals: event-*timex* (5% de reducció relativa mitjana de l'error, 0.68 F1), event-DCT (3% de reducció relativa mitjana de l'error, 0.82 F1), esdeveniments principals (3% de reducció relativa mitjana de l'error, 0.61 F1), i esdeveniments subordinats (2% de reducció relativa mitjana de l'error, 0.66 F1). La classificació d'esdeveniments és l'única tasca que no ha sigut millorada pels trets semàntics (0% de reducció relativa mitjana de l'error, 0,79 F1), però s'hi han proposat mètodes alternatius per a prendre avantatge de la semàntica en aquesta tasca.
- *S'hi obté un alt rendiment.* En comparació amb l'estat de la qüestió (SOA) per a l'anglès, TIPSem obté un resultat molt competitiu en totes les subtasques, la qual cosa assenyala que el nostre enfocament semàntic ofereix un alt rendiment. En particular, TIPSem supera el rendiment del SOA en la tasca de processament d'esdeveniments (+0,05 F1), fet que representa una de les principals aportacions d'aquesta tesi (0,85 F1 en el processament d'esdeveniments).
- *La nostra aproximació és útil per a diferents llengües.* Les conclusions descrites han sigut confirmades en tots els idiomes inclosos en l'avaluació multilingüe: espanyol, italià i xinès. TIPSem supera el baseline i el rendiment del SOA en aquests idiomes. A més, fins on nosaltres

sabem, TIPSem és la primera aproximació al processament d'esdeveniments per a l'espanyol i el xinès.

Per tal d'avaluar extrínsecament TIPSem, s'ha aplicat per a crear una interfície gràfica innovadora per als documents textuais basada en la dimensió temporal (Time-Surfer). L'objectiu del Temps-Surfer és permetre als usuaris cercar i explorar la informació relativa a un període específic, esdeveniment o participant d'esdeveniment en un document.

A diferència de les actuals interfícies gràfiques, Time-Surfer no només es basa en la detecció automàtica de *timex*, sinó també dels esdeveniments i les relacions temporals. A través d'una interfície de línia de temps, proporciona als usuaris una imatge dinàmica de la distribució temporal dels esdeveniments d'un document. Temps-Surfer ha sigut aplicat amb èxit en articles d'història i biogràfics de la Viquipèdia.

Time-Surfer s'ha avaluat mitjançant una avaluació orientada a l'usuari. S'ha mesurat el temps que tarden els usuaris a respondre preguntes temporals i el grau de satisfacció amb el Time-Surfer en comparació amb les interfícies clàssiques d'accés a la informació, en particular, la Viquipèdia.

Els resultats de l'avaluació mostren que, amb un temps curt d'adaptació, un usuari humà pot aprofitar la utilització del Time-Surfer per a explorar la informació temporal. D'altra banda, el grau de satisfacció obtinguts mostren que el Time-Surfer és útil, fàcil d'usar i aplicable per als individus avaluats.

Tenint en compte que cada error de TIPSem en el processament de la informació temporal es propaga a Time-Surfer, els resultats obtinguts demostren que el rendiment TIPSem sobre text general (històric i biogràfic), és prou alt per a servir d'entrada per a aquesta aplicació.

Finalment, en l'avaluació de TIPSem, concloem que cal més coneixement. Pel que fa als errors que TIPSem deixa sense resoldre, a més dels procedents de les eines de PLN utilitzades o errors humans d'anotació, aquests són casos que requereixen una anàlisi lingüística en un nivell superior a la semàntica. Per exemple, alguns esdeveniments nominals i *timex* requereixen pragmàtica, i algunes *timex* requereixen la resolució de l'anàfora temporal. A més, les relacions temporals sovint requereixen un raonament causal, l'anàlisi del discurs i coneixement del món.

6.2 Treball futur

Proposem investigacions futures en diverses línies.

- *Millorar el TIPSem*: Una clara limitació que trobem en el nostre enfocament semàntic és la classificació d'esdeveniments. Els trets semàntics aplicats no són discriminatòris per a la classificació dels esdeveniments en la forma en què s'han definit. En aquesta línia, es proposa la inclusió dels sentits de les paraules mitjançant l'ús de tècniques de desambiguació del sentit, i també considerar les dependències introracionals sintàctiques i semàntiques entre els esdeveniments.

Utilitzant l'enfocament TIPSem, la majoria de les tasques es poden resoldre amb un rendiment acceptable, per damunt del 80%. No obstant això, sobrepassar aquesta barrera és complicat, ja que es requereix, per exemple, la resolució de l'anàfora, la capacitat de processament del discurs i el raonament de la causalitat. Això assenyala la necessitat de característiques de nivell pragmàtic. En aquesta línia, investigarem els recursos i els mètodes per a introduir aquest tipus de coneixement (per exemple, les relacions de causalitat de WordNet). Aquesta línia es pot explorar més enllà usant ontologies com SUMO¹.

Donat que TIPSem es basa en l'aprenentatge automàtic, depèn de les dades disponibles. Tenim la intenció de millorar-ne el rendiment aplicant l'aprenentatge actiu. TIPSem s'aplicarà sobre textos rellevants per a altres aplicacions; els errors detectats seran corregits i les correccions seran utilitzades per reentrenar el sistema.

- *Millorar les dades TimeML*: Donat que el processament de la informació temporal s'investiga a partir d'una estratègia basada en corpus, les conclusions presentades depenen de les dades. En aquesta tesi s'ha utilitzat l'actual estàndard d'or: el corpus del TempEval-2. No obstant això, la mida d'aquests corpus (al voltant de setanta mil paraules per a cada idioma) és relativament petita en comparació amb els conjunts de dades disponibles per a tasques més madures com el marcatge de PoS (1 milió de paraules). La mida de les dades anotades és rellevant per a l'obtenció de conclusions més fermes i un rendiment més alt usant tècniques d'aprenentatge automàtic. En concret, per a analitzar el rendiment del nostre enfocament semàntic, tant el contingut com la mida són importants.

D'una banda, per tal d'avaluar la contribució dels enfocaments semàntics, el contingut ha d'incloure alguns casos ambigus en el nivell morfosintàctic. Hem trobat un percentatge d'aquests casos en els corpus actuals per als diferents idiomes i n'esperaríem trobar una proporció

¹Suggested Upper Merged Ontology

similar en mostres més grans, però això només es pot demostrar en una nova avaluació sobre un conjunt de dades més gran.

D'altra banda, la mida relativament petita del corpus podria haver produït els següents efectes secundaris en la nostra avaluació. La mostra de dades d'entrenament podria no haver estat prou gran per a conèixer tots els elements lèxics que participen en entitats temporals (per exemple, totes les unitats temporals: *micro-segons*, i el mateix per als esdeveniments nominals). Això fa que la semàntica lèxica i els rols semàntics siguin més útils, perquè poden generalitzar les propietats de les entitats temporals i aprendre millor els models. Per exemple, totes les unitats temporals mostren una relació amb la classe semàntica *temps* de WordNet i, probablement, apareixen exercint un rol semàntic temporal, i anàlogament el mateix per a esdeveniments nominal. No obstant això, si les dades foren més grans, els models morfosintàctics oferirien una cobertura més ampla i l'avantatge de la semàntica seria més limitat. Per tant, per a avaluar el percentatge de la millora obtinguda que es veu afectat per aquest efecte secundari necessitem dur a terme una nova avaluació en un conjunt de dades més gran.

Per això, el corpus ha de ser ampliat i els mètodes han de ser reentrenats i reavaluats en un experiment més gran. Aquest és precisament l'objectiu del proper exercici d'avaluació internacional sobre el tractament de la informació temporal (TempEval-3)², en què es farà servir un corpus d'un milió de paraules. Com que l'anotació humana és costosa, part de l'esforç d'anotació es durà a terme de manera semiautomàtica amb la combinació dels tres millors sistemes de l'estat de la qüestió: TTK, TRIOS, i el nostre enfocament semàntic: TIPSem. En aquesta línia, serem part de l'organització TempEval-3.

- *Millorar el Time-Surfer*: La principal limitació de Time-Surfer, a més de la seua dependència de l'eficàcia de TIPSem, és la seua dependència de l'eficiència de TIPSem. Cal un processament substancial per anotar la informació temporal de cada document, la qual cosa requereix preprocessament *offline*. L'eficiència en el processament de la informació temporal és un fet encara inexplorat en el camp i no s'ha mesurat en els exercicis d'avaluació. No obstant això, en moltes aplicacions reals, com Time-Surfer, el temps de resposta és crucial. Un bon rendiment motiva els consumidors potencials a aplicar els resultats de la

²<http://www.cs.york.ac.uk/semeval/>

investigació en els seus negocis.

A més, per a convertir el Time-Surfer en una aplicació rendible, cal una avaluació centrada en una tasca concreta per a entendre millor les fortaleces i debilitats de la interfície i recollir noves idees per a refinar la seua aplicació en determinats sectors professionals. En aquesta línia, l'educació i el periodisme s'analitzaran com a professions amb usuaris potencials de Time-Surfer.

- *Aplicar TIPSem a la cerca de respostes*: Prendrem com a punt de partida les nostres investigacions anteriors en cerca de respostes realitzades amb un enfocament limitat al processament de *timex*. L'enfocament serà comparat amb el nostre enfocament de processament TimeML per a analitzar els avantatges de tenir en compte també els esdeveniments i les relacions temporals en la cerca de respostes temporals complexes.

6.3 Producció científica

Relacionada amb la tesi

La llista següent inclou la producció científica obtinguda com a resultat del desenvolupament d'aquesta tesi en ordre cronològic.

- (Llorens et al., 2009). **Temporal Expression Identification based on Semantic Roles (NLDB'09)**. Una anàlisi preliminar de la contribució dels rols semàntics de reconeixement d'expressions temporals en anglès amb un enfocament basat en regles.
- (Llorens et al., 2009b). **From Semantic Roles to Temporal Information Representation (MICAI'09)**. L'extensió dels treballs previs a la comptabilitat de llengua espanyola per a provar la independència del llenguatge dels beneficis de la semàntica.
- (Llorens et al., 2009c). **Using Semantic Networks to Identify Temporal Expressions from Semantic Roles (RANLP'09)**. La inclusió de les xarxes semàntiques (WordNet) en l'enfocament multilingüe anterior per al reconeixement de *timex*. S'ha demostrat que els dos tipus de semàntica ofereixen resultats complementaris i, per tant, la seua combinació és útil.
- (Llorens et al., 2009a). **TimeML Temporal Expressions Detection for Catalan Using Semantic Roles and Semantic Networks (SEPLN'09)**. L'extensió dels treballs previs al català.
- (Llorens et al., 2010a). **TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles (COLING'10)**. Aquest document suposa dos canvis fonamentals en els nostres treballs anteriors: (i) inclou

els esdeveniments en la nostra recerca, i (ii) canvia les regles per a l'aprenentatge automàtic. L'article inclou una avaluació detallada d'un enfocament basat en CRF per a reconèixer i classificar els esdeveniments TimeML. Els resultats van mostrar que la semàntica millora el tractament dels tipus més complexos d'esdeveniments (*i. e.*, substantiu, adjectiu).

- (Llorens et al., 2010b). **TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2 (SemEval'10, ACL)**. Aquest article descriu l'extensió del nostre enfocament basat en la semàntica a tots els elements TimeML. A més, suposa l'avaluació externa del nostre enfocament en el context d'un exercici d'avaluació internacional important (SemEval). Els resultats confirmen els avantatges de la semàntica i l'alt rendiment del nostre enfocament (TIPSem) en comparació amb la resta dels enfocaments d'estat de la tècnica. Per a l'anglès, TIPSem va obtenir el primer lloc en el processament d'esdeveniments i la classificació de relacions event-DCT. En la resta de tasques, s'obté la segona o la tercera posició. Per a l'espanyol, TIPSem va obtenir el primer lloc en totes les tasques. A més, TIPSem era l'únic sistema multilingüe.
- (Llorens et al., 2011b). **Time-Surfer: Time-based Graphical Access to Document Content (ECIR'11)**. Presenta l'aplicació que es descriu en aquesta tesi: Time-Surfer. L'article va ser preparat en col·laboració amb el professor Robert Gaizauskas.
- (Llorens, Saquete, & Navarro-Colorado, 2011). **Applying Semantic Knowledge to the Automatic Processing of Temporal Expressions and Events in Natural Language (Journal IPM, submitted: Dec. 2010)**. Una anàlisi detallada dels avantatges de l'aplicació de la semàntica per al processament de *time*x i esdeveniments (reconeixement i classificació).
- (Llorens et al., 2011c). **Data-Driven Approach Based on Semantic Roles for Recognizing Temporal Expressions and Events in Chinese (NLDB'11, Accepted)**. Una extensió de l'enfocament TIPSem al reconeixement de *time*x i d'esdeveniments, desenvolupada en col·laboració amb la Universitat de Chongqing (Xina).
- (Llorens et al., 2011a). **Syntax-Motivated Context Windows of Morpho-Lexical Features for Recognizing Time and Event Expressions in Natural Language (NLDB'11, Accepted)**. Un estudi detallat del context morfosintàctic per a fer front al reconeixement de *time*x i esdeveniments.
- (Caselli et al., 2011). **Data-Driven Approach Using Semantics for Recognizing and Classifying TimeML Events in Italian (RANLP'11, Accepted)**. Una extensió del mètode de processament d'esdeveniments TIPSem en italià, en col·laboració amb l'ILC-CNR de Pisa (Itàlia).

Derivada d'altres investigacions

La llista següent inclou la producció científica obtinguda com a resultat de la col·laboració en el desenvolupament d'altres treballs de recerca en PLN i ciències de la computació en general (ordre cronològic invers).

- (Lloret, Llorens, Moreda, Saquete, & Palomar, 2011). **Text Summarization Contribution to Semantic Question Answering: New Approaches for Finding Answers on the Web** (Journal IJIS'11).
- (Moreda et al., 2011). **Combining Semantic Information in Question Answering Systems** (Journal IPM'10).
- (Saquete et al., 2009). **Enhancing QA Systems with Complex Temporal Question Processing Capabilities** (Journal JAIR'09).
- (Pardiño, Gómez, Llorens, Muñoz-Terol, Navarro-Colorado, Saquete, Martínez-Barco, Moreda, & Palomar, 2009). **IBQAst: A Question Answering System for Text Transcriptions** (CLEF'08).
- (Pardiño, Gómez, Llorens, Muñoz, Navarro, Saquete, Martinez-Barco, Moreda, & Palomar, 2008). **Adapting IBQAS to work with text transcriptions in QAst Task: IBQAst** (CLEF'08).
- (Moreda, Llorens, Saquete, & Palomar, 2008a). **Automatic Generalization of a QA Answer Extraction Module Based on Semantic Roles** (IBERAMIA'08).
- (Moreda, Llorens, Saquete, & Palomar, 2008c). **Two Proposals of a QA Answer Extraction Module Based on Semantic Roles** (MICAI'08).
- (Moreda, Llorens, Saquete, & Palomar, 2008b). **The influence of semantic roles in QA: a comparative analysis** (SEPLN'08).
- (Llorens & Saquete, 2008). **Búsqueda de respuestas basada en conocimiento semántico (document tècnic)**.
- (Restrepo, Llorens, & Cuenca, 2008). **Implementación en FPGA de una cámara IP** (JCRA'08).
- (Cuenca, Ramos, Llorens, & Macia, 2008). **Reconfigurable Architecture for Embedding Web Services** (IEEE SPL'08).
- (Cuenca, Grediaga, Llorens, & Albero, 2007a). **Performance Evaluation of FPGA-Embedded Web Servers** (IEEE ICECS'07).
- (Cuenca, Llorens, & Albero, 2007b). **Evaluación de servidores web embebidos sobre FPGAs** (JCRA-ICECS'07).
- (Llorens, Ivars, Buiges, Cabo, Climent, & Fuster, 2005). **DomoSim, Simulador de sistemas domóticos** (EID'05).

6.4 Producció de programari

El desenvolupament d'aquesta tesi va generar dues aplicacions informàtiques: TIPSem i Time-Surfer.

TIPSem

TIPSem és un programari per al processament multilingüe de la informació temporal dels textos en llenguatge natural. En l'actualitat, només les versions en anglès i espanyol estan integrats en el paquet. L'italià i el xinès s'inclouran en futures versions de programari TIPSem. Una demostració del sistema de TIPSem està disponible en línia³. El TIPSem ha sigut inscrit en el Registre de la Propietat Intel·lectual espanyol (codi A_730_10).

Time-Surfer

Time-Surfer és una interfície web gràfica per a explorar i buscar en el contingut dels documents centrada en la informació temporal. La interfície requereix una eina de processament de la informació temporal per a anotar els documents d'entrada en l'estàndard TimeML. És independent del llenguatge. Atès que s'usa TIPSem (A-730-10) com a eina de processament de la informació temporal, es limita actualment a anglès i espanyol. Una demostració del sistema Time-Surfer està disponible en línia⁴. El Time-Surfer ha sigut inscrit en el Registre de la Propietat Intel·lectual espanyol (codi A_300_11).

³<http://gplsi.dlsi.ua.es/demos/TIMEE/>

⁴<http://gplsi.dlsi.ua.es/demos/TIMEE/Time-Surfer/>

Apèndix A

TIMEK: Base de coneixements temporals

En aquest apèndix es descriuen el recurs TIMEK. TIMEK és una base de coneixements temporals utilitzada per TIPSem. Aquest recurs té dos objectius principals: (i) proporcionar el coneixement temporal necessari per a obtenir algunes característiques per a la classificació i la normalització de *timex*, i (ii) proporcionar a TIPSem un mètode per a la normalització de *timex* en ISO 8601.

TIMEK es pot dividir en dues parts principals:

- Coneixement temporal: el coneixement sobre els conceptes temporals, com ara les unitats temporals (p. e.: *anys, dies, minuts, etc.*), els mesos de l'any (*gener, febrer, etc.*), els dies de la setmana (p. e.: *dilluns, dimarts, etc.*), i així successivament.
- Conjunts de regles: els conjunts de regles per a l'obtenció de característiques i per a l'obtenció de valors normalitzats en ISO8601.

Les seccions següents detallen els coneixements i les regles incloses en TIMEK per a l'anglès. TIMEK també inclou aquestes mateixes regles per a l'espanyol, i en el futur s'ampliarà a altres idiomes.

Coneixement temporal

Aquesta part de TIMEK es compon de diferents patrons (expressions regulars) per a identificar els elements de les expressions temporals. Aquests elements són els següents:

- **Unitats temporals (TUnit):** Les unitats temporals incloses es representen en l'expressió regular mostrada en (136).

(136) `(seconds|minute|hour|day|week|month|quarter|year|centur(y|ies)|millennium)s?`

- **Mesos de l'any:** Els mesos inclosos es representen en l'expressió regular que es mostra en (137).

(137) `((January|February|March|April|May|June|July|August|September|October|November|December)| (Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Sept|Oct|Nov|Dec)(\.)?)`

- **Dies de la setmana (DOW):** Els dies de la setmana inclosos es representen en l'expressió regular mostrada en (138).

(138) `(Monday|Tuesday|Wednesday|Thursday|Friday|Saturday|Sunday| (Mon|Tue|Wed|Thu|Fri|Sat|Sun)(\.)?)`

- **Dècades:** Les dècades incloses es representen en l'expressió regular que es mostra en (139).

(139) `(twenties|thirties|forties|fifties|sixties|seventies|eighties|nineties)`

- **Hores del dia (TOD):** Les hores del dia incloses es representen en l'expressió regular que es mostra en (140).

(140) `(morning|afternoon|evening|night|midnight)`

- **Estacions:** Les estacions incloses es representen en l'expressió regular que es mostra en (141).

(141) `(spring|summer|(autumn|fall)|winter)`

- **Granularitat d'hores:** Les granularitats d'hores es representen en l'expressió regular que es mostra en (142).

(142) `(?i)(seconds|minute(s)?|hour(s)?|"+TOD+")`

- **Indicadors de repetició:** Els indicadors de repetició inclosos es representen en l'expressió regular que es mostra en (143).

(143) `(?i)(each(\s|_)+.*|every(\s|_)+.*|.*"+DOW"+s.*|
(hour|day|week|month|quarter|year)ly)`

- **Nombres:** Els nombres escrits (p. e.: *one, eighty, two hundred and twenty one*) són traduïts en els seus valors per la funció NUMEK – veure (144).

(144) `one → 1
eighty → 80
two hundred and twenty one → 221`

Conjunts de regles

Les regles definides s'utilitzen per a obtenir el valor d'algunes de les característiques de TIPSem i per a normalitzar les *timex* a l'estàndard ISO 8601.

Obtenció de característiques

- **Patró (Pattern):** Donada una *timex*, un patró es construeix de la següent manera. En primer lloc, el nombre escrit es converteixen al seu valor numèric (per exemple, tres dies → 3 dies). En segon lloc, els següents elements de les *timex* es tradueixen:

- Unitats temporals (p. e.: *years, months, minute*) → TUNIT
- Nombres → NUM
- Mesos (p. e.: *June, December*) → MONTH
- Dies de la setmana (p. e.: *Monday, Saturday*) → WEEKDAY

L'exemple (145) mostra algunes *timex* traduïdes a patrons.

(145) a. next Monday → next WEEKDAY
b. June 1999 → MONTH NUM
c. two weeks ago → NUM TUNIT ago

- **Granularitat d'hora (Time granularity):** Una característica booleana que s'estableix a cert si la *timex* té granularitat d'hores. Aquesta captura si una data s'especifica en unitats de temps és més petita que els dies (hores, minuts i segons, etc.). Aquesta característica s'utilitza en la classificació de *timex*. És útil distingir entre les dates i les hores.
- **Indicador de repetició (Set indicator):** Una característica booleana activada quan la *timex* conté indicadors de repetició. Aquesta característica s'utilitza en la classificació de *timex*. És útil per a distingir conjunts (repeticions) de la resta dels tipus de *timex*.

Normalització ISO 8601

El procés de normalització en TIPSem consisteix en dos passos: (i) obtenir el tipus de normalització, i (ii) l'aplicació d'una funció de normalització, segons el tipus de normalització.

El recurs TIMEK conté les regles de normalització requerides pel segon pas. Les regles relatives a cada tipus de normalització són les següents:

- **ISO_explicit:** Aquestes regles s'utilitzen per a normalitzar *timex* explícites i completes, és a dir, les que inclouen informació explícita i completa sobre la seua ubicació en el temps (per exemple, maig de 2001). Se segueixen els passos següents per a normalitzar els diferents tipus de *timex*:
 - Si l'expressió ja està en format ISO8601 (per exemple, 1999-02-01), ja està normalitzada.
 - En cas contrari, cal detectar si es tracta d'una de les següents granularitats, i calcular la normalització corresponent:
 - * segles (p. e.: *20th century* → 19)
 - * dècades (p. e.: *eighties* → 198)
 - * anys (p. e.: *nineteen eighty-three* → 1983)
 - * mesos (p. e.: *June 1983* → 1983-06)
 - * dies (p. e.: *June 8th, 1983* → 1983-06-08)
 - * hores (p. e.: *June 8th, 1983 11 p.m.* → 1983-06-08T23:00)
- **ISO_function:** Aquestes normes s'utilitzen per a normalitzar *timex* implícites o incompletes, és a dir, aquelles que són relatives a la data

de creació del document (DCT) ¹. El sistema comprova si la seqüència de *timex* és un dels següents subtipus i, una vegada que es determina, aplica les regles corresponents. Els subtipus i les regles associades són:

- **Diferència explícita de la DCT** (p. e.: *yesterday, two months ago, etc.*): es calcula la diferència temporal respecte a la DCT per normalitzar la *timex* (p. e.: *yesterday* → DCT - 1 day; *two months ago* → DCT - 2 months).
 - **Referència especificada** (p. e.: *next week, last Monday*): Utilitzant la DCT i el modificador (p. e.: last, past, previous, next, etc.), la data explícita corresponent es busca al calendari (p. e.: next week → current week + 1; last Monday → DCT - [1,7] days – depending on the current day: *Monday, Tuesday, etc.*).
 - **Referència no especificada** (p. e.: *October, Monday*): Usant la DCT i el temps verbal del verb que governa la *timex*, la data explícita corresponent es busca al calendari (p. e.: *October with a past tense* → *closest past October in the calendar*).
- **Període:** Aquestes regles s'utilitzen per a normalitzar *timex* que representen durades. Les expressions període segueixen el patró: "Number TUnit(Number TUnit)*"(p. e.: *two days and six hours*). Les regles de TIMEK per a aquest patró i de forma recursiva guarda els TUnits associats amb els números en una matriu. Finalment, la normalització es duu a terme seguint l'algorisme que es mostra en (146), on la matriu Tunit-nombre és *elements*.

(146) Normalize Period Algorithm

```

write "P";

granularity = date;
for (i=0;i<elements.size;i++){
    if(elements[i][TUnit].matches(hour|minute|second)
        && granularity==date){
        granularity = time;
        write "T";
    }
    write elements[i][number];
    write substring(elements[i][TUnit],0,1);
}

```

¹La *timex* pot ser relativa a altres *timex* també, però la versió actual del nostre enfocament no les té en compte.

Seguint l'algorisme *two days and six hours* es normalitza com a *P2DT6H*.

- **ISO_set:** Aquestes regles s'usen per a normalitzar *timex* que representen repeticions (patrons recurrents en el temps). TIMEK només inclou regles per a normalitzar els següents tipus de normalització depenent de la seua granularitat:
 - **Anys:** XXXX (p. e.: *yearly, each year, every year*)
 - **Mesos:** XXXX-XX (p. e.: *monthly, each month, every month*)
 - **Setmanes:** XXXX-WXX (p. e.: *weekly, each week, every week, each Monday*)
 - **Dies:** XXXX-XX-XX (p. e.: *daily, each day, every day*)
- **Present_ref, past_ref i future_ref:** Aquests tipus no tenen funcions de normalització pel fet que aquests ja són els valors ISO 8601 per a referir-se a una *timex* difusa com *actualment, fa alguns anys, o en el futur*.

Apèndix B

Comparació de resultats CRF-SVM

En aquest apèndix s'ofereix informació detallada sobre el rendiment obtingut en les diferents subtasques de processament d'informació temporal aplicant camps aleatoris condicionals (CRF) i màquines de suport vectorial (SVM).

Per a CRF, s'ha utilitzat el CRF++ toolkit¹ i el procés d'aprenentatge es realitza mitjançant els paràmetres *one-vs-all* method, $C=1$ i polynomial-degree=2.

Per als SVM, s'ha utilitzat Yamcha² i el procés d'aprenentatge es realitza mitjançant els paràmetres: *one-vs-all* method, $C=1$ i polynomial-degree=2.

Les configuracions dels paràmetres descrits van ser seleccionades pel fet que van obtenir els millors resultats en una experimentació comparativa prèvia.

En les subseccions següents es mostren els resultats obtinguts per cadascuna de les tècniques amb les mateixes característiques (TIPSem) en les subtasques avaluades en el TempEval-2 en anglès.

Processament de *timex*

La taula B.1 mostra els resultats obtinguts per al processament d'expressions temporals.

Com es mostra en la taula, la variant de TIPSem-CRF supera la versió basada en SVM en el reconeixement en totes les mesures (precisió, cobertura

¹<http://crfpp.sourceforge.net/>

²<http://chasen.org/~taku/software/YamCha/>

| Approach | Recognition | | | Classification | Normalization |
|-------------------|-------------|--------|---------------|----------------|---------------|
| | precision | recall | $F_{\beta=1}$ | accuracy | accuracy |
| TIPSem-CRF | 0.93 | 0.84 | 0.88 | 0.95 | 0.78 |
| TIPSem-SVM | 0.86 | 0.83 | 0.85 | 0.96 | 0.80 |

Taula B.1: Processament de *timex* CRF *vs.* SVM (anglès)

i F1). No obstant això, en la classificació i normalització els SVM ofereixen uns resultats lleugerament millors.

Processament d'esdeveniments

La taula B.2 mostra els resultats obtinguts per al reconeixement d'esdeveniments.

| Approach | Recognition | | | Classification |
|-------------------|-------------|--------|---------------|----------------|
| | precision | recall | $F_{\beta=1}$ | accuracy |
| TIPSem-CRF | 0.82 | 0.87 | 0.85 | 0.80 |
| TIPSem-SVM | 0.78 | 0.86 | 0.82 | 0.80 |

Taula B.2: Processament d'esdeveniments CRF *vs.* SVM (anglès)

Una vegada més, la variant de TIPSem-CRF supera la versió basada en SVM en el reconeixement en totes les mesures (precisió, cobertura i F1), a més, en la classificació de la versió, SVM ofereix uns resultats lleugerament millors (no perceptible en dos punts de precisió decimal).

Processament de relacions temporals

La taula B.3 mostra els resultats obtinguts per a la classificació de les relacions temporals.

| Approach | timex-event | dct-event | main events | subordinated events |
|-------------------|-------------|-------------|-------------|---------------------|
| | accuracy | accuracy | accuracy | accuracy |
| TIPSem-CRF | 0.68 | 0.82 | 0.61 | 0.66 |
| TIPSem-SVM | 0.58 | 0.78 | 0.31 | 0.50 |

Taula B.3: Processament de relacions temporals CRF *vs.* SVM (anglès)

Finalment, en les diferents tasques de categorització, CRF ofereix millors resultats de precisió.

CRF *vs.* SVM: resum

Com es pot veure a través de les diferents taules, CRF supera SVM en el reconeixement de *timex* i esdeveniments. Per tant, els CRF se seleccionen en l'enfocament de TIPSem com a tècnica d'aprenentatge per a les tasques de reconeixement, que s'adapten perfectament a aquesta tècnica d'etiquetatge de seqüències.

En segon lloc, SVM supera CRF en la classificació de *timex* i la normalització, i la classificació d'esdeveniments. Per tant, aquestes tasques es duen a terme en l'enfocament TIPSem utilitzant classificadors SVM.

Finalment, CRF ofereix més exactitud que SVM en totes les subtasques de categorització de relacions. Llavors, els CRF són seleccionats per a acometre aquestes tasques.

La taula B.4 mostra un resum de les tècniques d'aprenentatge aplicades per TIPSem a cada subtasca de processament de la informació temporal.

| Task | Subtask | ML Technique |
|-------------------------------------|-----------------------|--------------|
| Timex Processing | Recognition | CRF |
| | Classification | SVM |
| | Normalization | SVM |
| Event Processing | Recognition | CRF |
| | Classification | SVM |
| Temporal Relation Processing | event-timex | CRF |
| | event-DCT | CRF |
| | main events | CRF |
| | subord. events | CRF |

Taula B.4: Selecció de tècniques d'ML en TIPSem

Apèndix C

Validació creuada i rellevància estadística

En aquest apèndix s'inclou una avaluació detallada, una validació creuada de 10 conjunts del nostre enfocament (TIPSem) i del nostre baseline (TIPSem-B) sobre les dades del TempEval-2 per a l'anglès. L'avaluació de 10 conjunts dona un resultat de rendiment més fiable dels enfocaments, ja que és la mitjana de 10 experiments diferents en lloc d'un. D'altra banda, la contribució de la semàntica en el processament de la informació temporal és avaluada en 10 experiments emparellats –TIPSem-B (abans d'aplicar la semàntica) i TIPSem (després d'aplicar la semàntica)– sobre el qual s'aplica l'estadística inferencial per a demostrar la rellevància estadística de la diferència de les mitjanes en els resultats, de la millora mitjana, i de la reducció relativa de l'error mitjana.

Les dades oficials del TempEval-2 es divideixen en tres conjunts: un conjunt d'entrenament, un de test d'entitats temporals (*timex* i esdeveniments), i un de test de les relacions temporals (Event-Timex, Event-DCT, esdeveniments principals, i esdeveniments subordinats). Les dades de test d'entitats i relacions es van presentar de manera separada, ja que, en l'exercici d'avaluació TempEval-2, ambdós elements van ser avaluats en manera aïllada per a una millor captura de l'actuació dels diferents sistemes en el processament d'aquests elements diferents. L'exemple (147) mostra una breu descripció de la divisió de dades.

(147) TempEval-2 English dataset:

Training (entities and relations): 162 docs, 53K words

Test (entities): 9 docs, 5K words

Test (relations): 11 docs, 5K words

Tenim dos tipus d'elements (entitats i relacions), que es distribueixen aproximadament en un 90%/10% per a l'entrenament i test respectivament. És a dir, el 90% de les entitats per a l'entrenament i el 10% per al test, i el 90% de les relacions per a l'entrenament i el 10% per al test.

Per tal de dur a terme una validació creuada de 10 conjunts, hem generat nou conjunts complementaris 90%/10% per a tots dos elements per separat. Per a les entitats, tenint en compte que el primer conjunt (fold-1) és equivalent a l'original de TempEval-2, hem construït els 10 conjunts que es descriuen en la taula C.1. Anàlogament, per a les relacions, hem construït els 10 conjunts que es descriuen en la taula C.2.

| Dataset | Training | | Test | | test 1st file |
|----------------|----------|-------|------|-------|-----------------------|
| | docs | words | docs | words | |
| Fold-1 | 162 | 53450 | 9 | 4849 | APW19980306.1001 |
| Fold-2 | 161 | 52343 | 10 | 5956 | ABC19980108.1830.0711 |
| Fold-3 | 155 | 52741 | 16 | 5558 | APW19980227.0476 |
| Fold-4 | 156 | 52305 | 15 | 5994 | NYT19980424.0421 |
| Fold-5 | 139 | 52342 | 32 | 5957 | ed980111.1130.0089 |
| Fold-6 | 146 | 52428 | 25 | 5871 | wsj_0263 |
| Fold-7 | 164 | 52306 | 7 | 5993 | wsj_0570 |
| Fold-8 | 149 | 52053 | 22 | 6246 | wsj_0637 |
| Fold-9 | 154 | 52395 | 17 | 5904 | wsj_0786 |
| Fold-10 | 153 | 52328 | 18 | 5971 | wsj_0973 |

Taula C.1: 10 conjunts de dades (entitats anglès)

| Dataset | Training | | Test | | test 1st file |
|----------------|----------|-------|------|-------|-----------------------|
| | docs | words | docs | words | |
| Fold-1 | 162 | 53450 | 11 | 4764 | ABC19980120.1830.0957 |
| Fold-2 | 163 | 52258 | 10 | 5956 | ABC19980108.1830.0711 |
| Fold-3 | 157 | 52656 | 16 | 5558 | APW19980227.0476 |
| Fold-4 | 158 | 52220 | 15 | 5994 | NYT19980424.0421 |
| Fold-5 | 141 | 52257 | 32 | 5957 | ed980111.1130.0089 |
| Fold-6 | 148 | 52343 | 25 | 5871 | wsj_0263 |
| Fold-7 | 166 | 52221 | 7 | 5993 | wsj_0570 |
| Fold-8 | 151 | 51968 | 22 | 6246 | wsj_0637 |
| Fold-9 | 156 | 52310 | 17 | 5904 | wsj_0786 |
| Fold-10 | 155 | 52243 | 18 | 5971 | wsj_0973 |

Taula C.2: 10 conjunts de dades (relacions anglès)

La tasca de processament d'informació temporal es pot dividir en subtasques representades en un dels dos conjunts de dades descrits:

- **Processament de *timex* (entitats):** reconeixement, classificació, i normalització de *timex*.
- **Processament d'esdeveniments (entitats):** reconeixement i classificació d'esdeveniments.
- **Processament de relacions temporals (relacions):** categorització de relacions event-*timex*, event-DCT, esdeveniments principals, i esdeveniments subordinats.

Per a cadascuna de les subtasques avaluades en el conjunt de dades corresponent, les subseccions següents reporten precisió (P), cobertura (R), i F1 per a tasques de reconeixement, i exactitud¹ per a les tasques de classificació, normalització i categorització. A més dels resultats de TIPSem i TIPSem-B, s'inclouen les seues diferències de puntuació i el resultat mitjà obtingut de la mitjana dels 10 conjunts.

Finalment, s'inclou la significació estadística de: les diferències de puntuació, la millora i la reducció relativa d'errors. S'utilitza el t-test de Student per provar si les diferències en els resultats són significatives o s'obtenen per casualitat. El t-test avalua si les mitjanes de dos grups són estadísticament diferents entre si. Utilitzem la prova t per mostrar si l'aplicació de la semàntica introdueix una contribució real a les diferents tasques. En aquest cas, es defineix un experiment emparellat amb el resultat abans d'aplicar la semàntica (TIPSem-B) i el resultat després d'aplicar la semàntica (TIPSem) calculat sobre les mateixes dades. En l'escenari descrit, hi ha 10 conjunts d'entrenament/test i, per tant, 10 experiments emparellats [TIPSem-B, TIPSem].

El t-test emparellat de Student és apropiat, ja que tenim un experiment la sortida del qual és contínua (puntuació $\mathbb{R}[0, 1]$); les diferències de puntuació que obtenim mostren, en general, una distribució normal, i 10 és un nombre apropiat de parells per a l'aplicació de l'estadística inferencial. A més, en aquest experiment computacional controlat (les mateixes dades, mateixos algorismes i models), hi ha una influència gairebé nul·la dels factors externs, com els presents per als experiments, per exemple, en éssers humans.

¹Equivalent a complete accuracy descrita en la secció 4.1.

L'objectiu és demostrar que la hipòtesi nul·la (*Els resultats de TIPSem són iguals als de TIPSem-B*) és falsa. En particular, la hipòtesi alternativa ha d'indicar una direcció positiva:

$$\text{Rendiment de TIPSem} > \text{Rendiment de TIPSem-B}$$

Per això, s'aplica un t-test emparellat d'una sola cua (direcció positiva). La fórmula (D) mostra com es porta a terme la inferència estadística del t-test descrit.

$$Diff_{fold} = TIPSem_{fold} - TIPSemB_{fold} \quad (C.1)$$

$$\overline{Diff} = \sum_{fold=1}^n (Diff_{fold}) \quad (C.2)$$

$$S = \sqrt{\frac{\sum_{fold=1}^n (Diff_{fold} - \overline{Diff})^2}{n}} \quad (C.3)$$

$$df = n - 1 \quad (C.4)$$

$$t = \frac{\overline{Diff} - \mu_0}{\frac{S}{\sqrt{n}}} \quad (C.5)$$

$$p(one - tail) = \frac{Student.t.distribution(t, df)}{2} \quad (C.6)$$

Suposant que $Diff_{fold}$ són les diferències entre les puntuacions de TIPSem i TIPSem B, S la desviació estàndard, n com el nombre de parells (10), df els graus de llibertat (9), i establint μ_0 com la diferència nul·la entre TIPSem i TIPSem-B ($\mu_0 = 0$), el valor de t serà positiu si la mitjana de rendiment de TIPSem és més alta que la de TIPSem-B, i negatiu si és menor. A més, si el valor de t és positiu, per tal de determinar que és prou alt per a descartar la hipòtesi nul·la en favor de l'alternativa s'ha d'obtenir un valor p d'una cua mirant la distribució t de la taula de Student, tenint en compte els graus de llibertat (df) i dividint per dos.

El valor de p per a una cua ens diu si la diferència de rendiment causada per l'ús de semàntica és prou gran per a dir que no és probable que haja estat una troballa casual. Si p és inferior al llindar o nivell de risc (nivell alfa) triat per a la significació estadística, llavors la hipòtesi nul·la és rebutjada en favor de la hipòtesi alternativa. Com més baix és el nivell alfa, més fiables són els resultats de significació obtinguts. En els nostres experiments, hem tingut en compte el nivell alfa que es considera generalment en els experiments

científics per a una fiabilitat alta: 0,10 (90% de confiança), 0,05 (95% de confiança), 0,01 (99% de confiança) i 0.005 (99,5% de confiança).

En les subseccions següents es mostren els resultats dels 10 conjunts i les proves de significació per a cada subtasca, és a dir, el processament de *timex*, el processament d'esdeveniments i el processament de relacions temporals.

Processament de *timex*

La validació creuada de 10 conjunts per al processament de *timex* es realitza sobre les dades de les entitats anteriorment descrites. La tasca consta de tres subtasques: reconeixement, classificació i normalització.

Reconeixement de *timex*

La taula C.3 mostra els resultats obtinguts per al reconeixement de *timex*.

| Dataset | TIPSem-B | | | TIPSem | | | Difference | | |
|----------------|----------|------|------|--------|------|------|------------|-------|-------|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| Fold-1 | 0.89 | 0.68 | 0.77 | 0.93 | 0.84 | 0.88 | 0.04 | 0.15 | 0.11 |
| Fold-2 | 0.89 | 0.64 | 0.75 | 0.89 | 0.82 | 0.85 | 0.00 | 0.17 | 0.11 |
| Fold-3 | 0.95 | 0.60 | 0.74 | 0.94 | 0.79 | 0.86 | -0.02 | 0.19 | 0.12 |
| Fold-4 | 0.84 | 0.50 | 0.63 | 0.83 | 0.60 | 0.70 | -0.01 | 0.10 | 0.07 |
| Fold-5 | 0.97 | 0.83 | 0.89 | 0.95 | 0.85 | 0.89 | -0.02 | 0.02 | 0.00 |
| Fold-6 | 0.96 | 0.84 | 0.90 | 0.95 | 0.88 | 0.91 | -0.01 | 0.03 | 0.01 |
| Fold-7 | 0.95 | 0.65 | 0.77 | 0.93 | 0.80 | 0.86 | -0.02 | 0.15 | 0.09 |
| Fold-8 | 0.94 | 0.83 | 0.88 | 0.95 | 0.89 | 0.92 | 0.01 | 0.06 | 0.04 |
| Fold-9 | 0.97 | 0.78 | 0.87 | 0.96 | 0.87 | 0.92 | -0.01 | 0.09 | 0.05 |
| Fold-10 | 0.94 | 0.82 | 0.87 | 0.91 | 0.81 | 0.86 | -0.03 | -0.01 | -0.02 |
| Mean | 0.93 | 0.72 | 0.81 | 0.92 | 0.81 | 0.87 | -0.01 | 0.10 | 0.06 |

| Comparison | | | | |
|----------------------------|----|--------|-------|----------------|
| TIPSem/TIPSem-B | | Mean | S | t (p one-tail) |
| Difference | P | -0.01 | 0.02 | negative |
| | R | 0.10 | 0.07 | 4.68 (0.005) |
| | F1 | 0.06 | 0.05 | 3.96 (0.005) |
| Improvement % | P | -0.69 | 2.09 | negative |
| | R | 14.86 | 10.77 | 4.36 (0.005) |
| | F1 | 7.67 | 6.23 | 3.89 (0.005) |
| Relative error reduction % | P | -21.58 | 31.25 | negative |
| | R | 31.47 | 17.57 | 5.66 (0.005) |
| | F1 | 26.30 | 19.90 | 4.18 (0.005) |

Taula C.3: Reconeixement de *timex* 10-fold (anglès)

Com es mostra en la taula, TIPSem obté uns valors mitjans alts de precisió (0,92), cobertura (0,81) i F1 (0,87). A més, la cobertura mostra una diferència positiva mitjana de +0,10 i, encara que la precisió mostra una petita diferència negativa (-0,01), la diferència final de F1 és positiva (+6 punts). Aquesta diferència és estadísticament significativa amb un nivell de confiança molt alt (99,5%). També la millora i la reducció relativa d'errors que TIPSem obté sobre TIPSem-B són significatives.

Classificació de *timex*

La taula C.4 mostra els resultats obtinguts per a la tasca de classificació de *timex*.

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|---------|----------------------|--------------------|------------------------|
| Fold-1 | 0.93 | 0.98 | 0.05 |
| Fold-2 | 0.87 | 0.89 | 0.03 |
| Fold-3 | 0.89 | 0.91 | 0.02 |
| Fold-4 | 0.90 | 0.95 | 0.04 |
| Fold-5 | 0.90 | 0.89 | -0.02 |
| Fold-6 | 0.92 | 0.92 | 0.01 |
| Fold-7 | 0.95 | 0.95 | 0.01 |
| Fold-8 | 0.95 | 0.96 | 0.01 |
| Fold-9 | 0.95 | 0.95 | 0.01 |
| Fold-10 | 0.96 | 0.96 | 0.00 |
| Mean | 0.92 | 0.94 | 0.02 |

| Comparison TIPSem/TIPSem-B | | Mean | S | t (p one-tail) |
|-------------------------------|-----|-------|-------|----------------|
| Difference | ACC | 0.02 | 0.02 | 2.65 (0.025) |
| Improvement % | ACC | 1.70 | 2.03 | 2.65 (0.025) |
| Relative error reduction % | ACC | 18.89 | 21.28 | 2.81 (0.025) |

Taula C.4: Classificació de *timex* 10-fold (anglès)

Com es mostra en la taula, TIPSem obté una exactitud mitjana alta (0,94), amb una diferència mitjana positiva de 0,02 sobre el baseline. Malgrat que aquesta diferència és menor que l'obtinguda per al reconeixement de *timex*, és estadísticament significativa amb un nivell de confiança alt (97,5%).

Normalització de *timex*

La taula C.5 mostra els resultats obtinguts.

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|---------|----------------------|--------------------|------------------------|
| Fold-1 | 0.75 | 0.81 | 0.06 |
| Fold-2 | 0.62 | 0.72 | 0.11 |
| Fold-3 | 0.62 | 0.71 | 0.09 |
| Fold-4 | 0.70 | 0.77 | 0.07 |
| Fold-5 | 0.55 | 0.58 | 0.03 |
| Fold-6 | 0.57 | 0.62 | 0.05 |
| Fold-7 | 0.74 | 0.78 | 0.05 |
| Fold-8 | 0.64 | 0.75 | 0.11 |
| Fold-9 | 0.61 | 0.66 | 0.05 |
| Fold-10 | 0.58 | 0.64 | 0.05 |
| Mean | 0.64 | 0.70 | 0.07 |

| Comparison | | Mean | S | t (p one-tail) |
|----------------------------|-----|-------|------|----------------|
| TIPSem/TIPSem-B | | | | |
| Difference | ACC | 0.07 | 0.03 | 8.33 (0.005) |
| Improvement % | ACC | 10.43 | 4.00 | 8.25 (0.005) |
| Relative error reduction % | ACC | 19.05 | 7.47 | 8.06 (0.005) |

Taula C.5: Normalització de *timex* 10-fold (anglès)

TIPSem obté una exactitud mitjana de 0.70, amb una diferència mitjana positiva de 0,07 sobre el baseline. Aquesta diferència és estadísticament significativa amb un nivell de confiança molt alt (99,5%).

Processament d'esdeveniments

El processament d'esdeveniments en la validació creuada de 10 conjunts es realitza sobre les dades de les entitats anteriorment descrites. Aquest consisteix en dues subtasques: reconeixement i classificació.

Reconeixement d'esdeveniments

La taula C.6 mostra els resultats obtinguts.

| Dataset | TIPSem-B | | | TIPSem | | | Difference | | |
|----------------|----------|------|------|--------|------|------|------------|------|------|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| Fold-1 | 0.82 | 0.80 | 0.81 | 0.82 | 0.87 | 0.85 | 0.00 | 0.07 | 0.03 |
| Fold-2 | 0.91 | 0.77 | 0.83 | 0.89 | 0.83 | 0.86 | -0.02 | 0.06 | 0.03 |
| Fold-3 | 0.89 | 0.77 | 0.83 | 0.88 | 0.83 | 0.85 | -0.01 | 0.06 | 0.03 |
| Fold-4 | 0.87 | 0.74 | 0.80 | 0.85 | 0.80 | 0.83 | -0.01 | 0.06 | 0.03 |
| Fold-5 | 0.90 | 0.77 | 0.83 | 0.90 | 0.82 | 0.85 | -0.01 | 0.05 | 0.02 |
| Fold-6 | 0.83 | 0.80 | 0.82 | 0.82 | 0.85 | 0.83 | -0.01 | 0.04 | 0.02 |
| Fold-7 | 0.80 | 0.79 | 0.80 | 0.78 | 0.83 | 0.81 | -0.02 | 0.04 | 0.01 |
| Fold-8 | 0.84 | 0.79 | 0.81 | 0.84 | 0.84 | 0.84 | 0.00 | 0.05 | 0.03 |
| Fold-9 | 0.89 | 0.82 | 0.85 | 0.87 | 0.85 | 0.86 | -0.02 | 0.02 | 0.01 |
| Fold-10 | 0.89 | 0.80 | 0.84 | 0.88 | 0.82 | 0.85 | -0.01 | 0.02 | 0.01 |
| Mean | 0.86 | 0.79 | 0.82 | 0.85 | 0.83 | 0.84 | -0.01 | 0.05 | 0.02 |

| Comparison | | | | |
|----------------------------|------|-------|----------------|---------------|
| TIPSem/TIPSem-B | Mean | S | t (p one-tail) | |
| Difference | P | -0.01 | 0.01 | negative |
| | R | 0.05 | 0.02 | 9.74 (0.005) |
| | F1 | 0.02 | 0.01 | 6.37 (0.005) |
| Improvement % | P | -1.33 | 0.81 | negative |
| | R | 6.17 | 2.07 | 9.42 (0.005) |
| | F1 | 2.45 | 1.23 | 6.29 (0.005) |
| Relative error reduction % | P | -9.25 | 5.47 | negative |
| | R | 22.26 | 6.79 | 10.36 (0.005) |
| | F1 | 11.13 | 5.26 | 6.69 (0.005) |

Taula C.6: Reconeixement d'esdeveniments 10-fold (anglès)

Com es mostra en la taula, TIPSem obté uns resultats mitjans alts de precisió (0,85), cobertura (0,83) i F1 (0,84). A més, la cobertura mostra una diferència positiva mitjana de +0,05 i, encara que la precisió mostra una petita diferència negativa (-0,01), la diferència final de F1 és positiva (+2 punts). Aquesta diferència, la millora, i la reducció relativa d'errors, són estadísticament significatives amb un nivell de confiança molt alt (99,5%).

Classificació d'esdeveniments

La taula C.7 mostra els resultats obtinguts per a la tasca de classificació d'esdeveniments.

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|---------|----------------------|--------------------|------------------------|
| Fold-1 | 0.80 | 0.78 | -0.02 |
| Fold-2 | 0.79 | 0.79 | 0.00 |
| Fold-3 | 0.75 | 0.74 | 0.00 |
| Fold-4 | 0.78 | 0.78 | 0.00 |
| Fold-5 | 0.74 | 0.74 | 0.00 |
| Fold-6 | 0.77 | 0.77 | 0.01 |
| Fold-7 | 0.76 | 0.76 | 0.00 |
| Fold-8 | 0.78 | 0.79 | 0.01 |
| Fold-9 | 0.81 | 0.81 | 0.00 |
| Fold-10 | 0.84 | 0.83 | -0.01 |
| Mean | 0.78 | 0.78 | 0.00 |

| Comparison TIPSem/TIPSem-B | | Mean | S | t (p one-tail) |
|-------------------------------|-----|-------|------|----------------|
| Difference | ACC | 0.00 | 0.01 | -0.24 (not-sg) |
| Improvement % | ACC | -0.05 | 0.84 | negative |
| Relative error reduction % | ACC | -0.43 | 3.23 | negative |

Taula C.7: Classificació d'esdeveniments 10-fold (anglès)

TIPSem obté una precisió mitjana de 0,83, amb una diferència mitjana nul·la sobre el baseline. De fet, la millora i la reducció d'errors són negatives. Aquesta diferència és gairebé nul·la i no significativa.

Això ajuda a confirmar els resultats obtinguts en la secció d'avaluació sobre el test de TempEval-2 (fold1). L'ús de la semàntica no introdueix una millora significativa en el rendiment del baseline. De fet, s'hi introdueix una lleugera disminució del rendiment. Per tant, l'ús de la semàntica en TIPSem no és eficient per a aquesta tasca.

Processament de relacions temporals

Les taules C.8, C.9, C.10 i C.11 mostren els resultats dels quatre tipus de relacions temporals.

Categorització de relacions Event-Timex

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|----------------|----------------------|--------------------|------------------------|
| Fold-1 | 0.65 | 0.68 | 0.03 |
| Fold-2 | 0.80 | 0.76 | -0.05 |
| Fold-3 | 0.72 | 0.80 | 0.08 |
| Fold-4 | 0.75 | 0.81 | 0.06 |
| Fold-5 | 0.54 | 0.55 | 0.01 |
| Fold-6 | 0.51 | 0.54 | 0.03 |
| Fold-7 | 0.64 | 0.63 | -0.01 |
| Fold-8 | 0.54 | 0.56 | 0.02 |
| Fold-9 | 0.60 | 0.60 | 0.00 |
| Fold-10 | 0.45 | 0.46 | 0.01 |
| Mean | 0.62 | 0.64 | 0.02 |

| Comparison | | Mean | S | t (p one-tail) |
|----------------------------|-----|------|-------|----------------|
| TIPSem/TIPSem-B | | | | |
| Difference | ACC | 0.02 | 0.03 | 1.61 (0.10) |
| Improvement % | ACC | 2.84 | 4.68 | 1.92 (0.05) |
| Relative error reduction % | ACC | 4.72 | 14.06 | 1.06 (not-sg) |

Taula C.8: Categorització Event-timex 10-fold (anglès)

TIPSem obté una exactitud mitjana de 0,64, amb una diferència mitjana positiva de +0,02 sobre el baseline. Aquesta diferència és estadísticament significativa amb un nivell de confiança del 90%. La millora obtinguda (2,84%) és significativa amb una confiança del 95%, però la reducció relativa d'errors (4,72) no és significativa.

Els resultats obtinguts per a relacions event-*timex* mostren que l'aplicació de la semàntica ofereix una diferència de mitjanes positiva més gran que 0 (0,02), però la significació d'aquesta diferència està en els límits dels nivells de confiança científica (90%). Per exemple, es pot observar que en el conjunt 2 (fold-2), l'aplicació de la semàntica resulta en una disminució del rendiment de -0,05.

Categorització de relacions Event-DCT

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|---------|----------------------|--------------------|------------------------|
| Fold-1 | 0.80 | 0.82 | 0.02 |
| Fold-2 | 0.75 | 0.74 | -0.01 |
| Fold-3 | 0.76 | 0.74 | -0.02 |
| Fold-4 | 0.64 | 0.64 | 0.00 |
| Fold-5 | 0.40 | 0.80 | 0.40 |
| Fold-6 | 0.80 | 0.82 | 0.02 |
| Fold-7 | 0.75 | 0.74 | -0.01 |
| Fold-8 | 0.76 | 0.74 | -0.02 |
| Fold-9 | 0.64 | 0.64 | 0.00 |
| Fold-10 | 0.40 | 0.80 | 0.40 |
| Mean | 0.67 | 0.75 | 0.08 |

| Comparison TIPSem/TIPSem-B | | Mean | S | t (p one-tail) |
|-------------------------------|-----|-------|-------|----------------|
| Difference | ACC | 0.08 | 0.16 | 1.51 (0.10) |
| Improvement % | ACC | 19.59 | 40.24 | 1.54 (0.10) |
| Relative error reduction % | ACC | 12.67 | 27.74 | 1.44 (0.10) |

Taula C.9: Categorització Event-DCT 10-fold (anglès)

TIPSem obté una exactitud mitjana de 0,75, amb una diferència mitjana positiva de +0,08 sobre el baseline. Aquesta diferència, la millora i la reducció d'errors són estadísticament significatives amb un nivell de confiança del 90%.

Cal destacar que per al cas de les relacions Event-DCT, els conjunts del 6 al 10 (principalment fixers WSJ) no contenen cap relació d'aquest tipus. Hem abordat aquest problema mitjançant l'ús dels conjunts de l'1 al 5 dues vegades. Aquestes dades no anotades ens impedeixen fer una avaluació de 10 conjunts normal. Això pot influenciar els resultats mitjans i de significació; per tant, els resultats presentats per a Event-DCT només donen una estimació indirecta.

Categorització de relacions d'esdeveniments principals

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|----------------|------------------------------|----------------------------|--------------------------------|
| Fold-1 | 0.58 | 0.61 | 0.03 |
| Fold-2 | 0.51 | 0.51 | 0.00 |
| Fold-3 | 0.45 | 0.48 | 0.02 |
| Fold-4 | 0.56 | 0.56 | 0.01 |
| Fold-5 | 0.53 | 0.54 | 0.01 |
| Fold-6 | 0.50 | 0.50 | 0.00 |
| Fold-7 | 0.48 | 0.52 | 0.04 |
| Fold-8 | 0.41 | 0.42 | 0.02 |
| Fold-9 | 0.42 | 0.44 | 0.02 |
| Fold-10 | 0.38 | 0.39 | 0.01 |
| Mean | 0.48 | 0.50 | 0.02 |

| Comparison | | | | | |
|-----------------------------------|------------|-------------|----------|-----------------------|--|
| TIPSem/TIPSem-B | | Mean | S | t (p one-tail) | |
| Difference | ACC | 0.02 | 0.01 | 3.80 (0.005) | |
| Improvement % | ACC | 3.27 | 2.63 | 3.93 (0.005) | |
| Relative error reduction % | ACC | 2.96 | 2.58 | 3.62 (0.005) | |

Taula C.10: Categorització d'esdeveniments principals 10-fold (anglès)

TIPSem obté una exactitud de 0,50, amb una diferència mitjana positiva de +0,02 sobre el baseline. Aquesta diferència, la millora, i la reducció relativa d'errors són estadísticament significatives, amb un nivell de confiança molt alt (99,5%).

Categorització de relacions d'esdeveniments subordinats

| Dataset | TIPSem-B accuracy | TIPSem accuracy | Difference accuracy |
|---------|----------------------|--------------------|------------------------|
| Fold-1 | 0.61 | 0.66 | 0.05 |
| Fold-2 | 0.50 | 0.51 | 0.01 |
| Fold-3 | 0.51 | 0.50 | -0.01 |
| Fold-4 | 0.48 | 0.48 | 0.00 |
| Fold-5 | 0.62 | 0.63 | 0.01 |
| Fold-6 | 0.70 | 0.70 | 0.00 |
| Fold-7 | 0.61 | 0.63 | 0.01 |
| Fold-8 | 0.37 | 0.37 | 0.00 |
| Fold-9 | 0.44 | 0.46 | 0.02 |
| Fold-10 | 0.44 | 0.44 | 0.00 |
| Mean | 0.53 | 0.54 | 0.01 |

| Comparison TIPSem/TIPSem-B | | Mean | S | t (p one-tail) |
|-------------------------------|-----|------|------|----------------|
| Difference | ACC | 0.01 | 0.01 | 2.07 (0.05) |
| Improvement % | ACC | 1.69 | 2.50 | 2.14 (0.05) |
| Relative error reduction % | ACC | 2.40 | 3.77 | 2.02 (0.05) |

Taula C.11: Categorització d'esdeveniments subordinats 10-fold (anglès)

TIPSem obté una exactitud mitjana de 0,55, amb una diferència mitjana positiva de +0,01 sobre el baseline. Aquesta diferència, la millora, i la reducció d'errors relatius són estadísticament significatives amb un nivell de confiança del 95%.

Resum de la validació creuada de 10 conjunts

En general, s’observa que l’ús de la semàntica (TIPSem) presenta un increment mitja estadísticament significatiu sobre el rendiment del baseline només basat en morfosintaxi (TIPSem-B). La taula C.12 mostra una versió resumida dels resultats, la reducció relativa mitjana de l’error entre TIPSem-B i TIPSem, i la significació de la diferència mitjana del rendiment.

| Task | Subtask | TIPSem mean F1 | Mean RER_S | Mean diff. SC |
|------------------------------|-------------------------------|-------------------|---------------|------------------|
| Timex | recognition F1 | 0.87 | 26% | 99.5% |
| | classification ACC | 0.94 | 19% | 97.5% |
| | normalization ACC | 0.70 | 19% | 99.5% |
| Event | recognition F1 | 0.84 | 11% | 99.5% |
| | classification ACC | 0.78 | 0% | - |
| Temporal Relation | event-<i>timex</i> ACC | 0.64 | 5% | 90% |
| | event-DCT ACC | 0.75 | 3% | 90% |
| | main events ACC | 0.50 | 3% | 99.5% |
| | subord. events ACC | 0.55 | 2% | 95% |

Taula C.12: Resum de la validació de 10 conjunts (RER_S: reducció relativa d’error per la semàntica, SC: fiabilitat de la significació)

Les tasques de reconeixement de *timex* i esdeveniments mostren una millora estable i significativa que afecta especialment la cobertura. La classificació i normalització de *timex* també obté una millora significativa en l’exactitud, mentre que la classificació d’esdeveniments no. Pel que fa a les tasques de categorització de relacions temporals, event-*timex* i event-DCT, s’observa una millora amb un nivell de confiança ajustat, mentre que les relacions d’esdeveniments principals i les de subordinació mostren una millora estadísticament significativa, amb un nivell de confiança més alt.

Val la pena esmentar que, a més d’aquest test de significació, es pot observar que en les tasques de reconeixement és difícil pensar en una millora fortuïta. Cal tenir en compte que les dades de test consten de cinc mil paraules, de les quals només 136 són part d’alguna *timex* i 498 són part d’esdeveniments. Si tenim en compte que el baseline ja troba el 80% dels casos, una millora necessita no modificar els elements ja trobats, i trobar els 20% (27 *timex* i 100 esdeveniments) restants en gairebé cinc mil paraules. Tenint en compte aquesta anàlisi, una modificació a l’atzar de l’anotació de les paraules anotades pel baseline té una baixa probabilitat de produir una millora. Aquesta és una raó addicional que suporta que la millora obtinguda mitjançant l’ús de la semàntica és significativa i no a causa

d'una coincidència. A més, contràriament als estudis en humans, en aquest experiment computacional controlat es pot garantir que la diferència dels resultats és produïda per l'aplicació de la semàntica i no per qualsevol altre factor extern.

En resum, la major part de les tasques es beneficien de la semàntica i en general, la contribució de la semàntica és significativa amb un alt nivell de confiança. La classificació d'esdeveniments és l'única que mostra la conclusió oposada, i això es deu al fet que la semàntica emprada és massa simple en comparació amb la complexitat d'aquesta tasca, com s'analitza en el capítol d'avaluació.

Apèndix D

Exercicis d'avaluació TempEval

S'han realitzat dos exercicis d'avaluació internacional de processament de la informació temporal TimeML: TempEval-1 i TempEval-2.

TempEval-1 (2007)

El primer TempEval, exercici d'avaluació TempEval-1 (Verhagen et al., 2007), es va dur a terme en el context de SemEval-2007, com una de les tasques. En TempEval-1 van ser avaluats diferents enfocaments de categorització automàtica de relacions temporals TimeML (TLINKs). L'exercici d'avaluació es va dividir en tres tasques (A, B i C). Una tasca es va dirigir a la categorització de les relacions temporals entre les expressions de temps i els esdeveniments que ocorren en la mateixa frase. La tasca B es va centrar en la categorització de les relacions temporals entre la data de la creació dels documents (DCT) i els esdeveniments. Finalment, la tasca C implicava la categorització de les relacions d'esdeveniment principals de les frases adjacents. Així, les relacions temporals definides en TimeML ([T, S, A] LINKS) es van simplificar a tres tipus diferents de TLINKs. D'altra banda, la tasca d'anotació es va centrar no en l'extracció, sinó en la categorització de TLINKs en sis tipus de relació: before, after, overlap, before-or-overlap, overlap-or-after i vague. Els enfocaments avaluats en aquest estudi estaven principalment basats en aprenentatge automàtic i característiques morfo-sintàctiques. La taula D.1 resumeix les puntuacions mitjanes i les millors F1 per a cada tasca.

En el TempEval-1, els sistemes més destacats que van aconseguir els mi-

| | timex-event | dct-event | main events |
|------------|--------------------|------------------|--------------------|
| Average F1 | 0.56 | 0.74 | 0.51 |
| Best F1 | 0.62 | 0.80 | 0.55 |

Taula D.1: TempEval-1 resultats F1

llors resultats són: WVALI (Puscasu, 2007), CU-TMP (Bethard & Martin, 2007) i LCC-TE (Min et al., 2007).

TempEval-2 (2010)

El TempEval-2 (2010) va avaluar el processament de *timex*, esdeveniment, i quatre tipus de relacions temporals; en l'actualitat l'avaluació és més important en el processament de la informació temporal. A més, en comptes d'avaluar només en anglès, es van crear corpus per a sis idiomes diferents. Aquesta tesi aborda una avaluació multilingüe de totes les entitats TimeML. Per aquesta raó, vam prendre l'avaluació TempEval-2 com a base per a les proves de la nostra proposta i per a comparar-la amb altres enfocaments.

En resum, l'avaluació TempEval-2 va cobrir l'anotació automàtica d'expressions temporals (tasca A), esdeveniments (tasca B), relacions Event-Timex (Tasca C¹), relacions Event-DCT (Tasca D²), les relacions d'esdeveniments principals (Tasca E³), i les relacions d'esdeveniments subordinats (Tasca F). Aquest exercici d'avaluació és un marc adequat per a comparar el rendiment dels sistemes esmentats en la tesi que van participar en la iniciativa. Una versió anterior del nostre enfocament, TIPSem, i del nostre baseline, TIPSem-B, van participar en aquesta avaluació (Llorens et al., 2010b).

La taula D.2 conté els resultats oficials del TempEval-2 per al processament d'expressions temporals (tasca A) en diferents idiomes. La taula D.3 conté els resultats oficials del TempEval-2 per al processament d'esdeveniments (tasca B) en diferents idiomes. Finalment, la taula D.4 mostra els resultats oficials del TempEval-2 per a les tasques de categorització de relacions (tasques C a F). En les taules, els millors resultats es destaquen en negreta i els sistemes estan ordenats per resultats F1 en ordre descendent.

Dins de TempEval-2, els millors resultats els han obtingut els sistemes HeidelTime, TIPSem, TRIOS, i NCSU. Tres sistemes van abordar totes les

¹Correspon a la tasca A del TempEval-1

²Correspon a la tasca B del TempEval-1

³Correspon a la tasca C del TempEval-1

tasques JU_CSE, TIPSem i TRIOS, i els dos últims van obtenir resultats competitius en general. Només el nostre sistema, TIPSem, va ser avaluat en més d'un idioma (anglès i espanyol). Finalment, no hi va haver participants en la resta d'idiomes: xinès, italià, francès i coreà.

Els resultats de TempEval-2 són lleugerament millors que els obtinguts en TempEval-1 en les tasques de categorització de relacions temporals.

| language | system | Recognition | | | Attributes | |
|----------|------------|-------------|-------------|-------------|-------------|-------------|
| | | precision | recall | F1 | type | value |
| English | HeidelTime | 0.90 | 0.82 | 0.86 | 0.96 | 0.85 |
| | TRIOS | 0.85 | 0.85 | 0.85 | 0.94 | 0.76 |
| | TIPSem | 0.92 | 0.80 | 0.85 | 0.92 | 0.65 |
| | KUL | 0.85 | 0.84 | 0.84 | 0.91 | 0.55 |
| | Edinburgh | 0.85 | 0.82 | 0.84 | 0.84 | 0.63 |
| | TERSEO | 0.76 | 0.66 | 0.71 | 0.98 | 0.65 |
| | TIPSem-B | 0.88 | 0.60 | 0.71 | 0.88 | 0.59 |
| | JU_CSE | 0.55 | 0.17 | 0.26 | 0.00 | 0.00 |
| Spanish | TIPSem | 0.95 | 0.87 | 0.91 | 0.91 | 0.78 |
| | UC3M | 0.90 | 0.87 | 0.88 | 0.91 | 0.83 |
| | TIPSem-B | 0.97 | 0.81 | 0.88 | 0.99 | 0.75 |

Taula D.2: TempEval-2 - resultats de processament de *timex*

| language | system | Recognition | | | attr. class |
|----------|---------------|-------------|-------------|-------------|----------------|
| | | precision | recall | F1 | |
| English | TIPSem | 0.81 | 0.86 | 0.83 | 0.79 |
| | TIPSem-B | 0.83 | 0.81 | 0.82 | 0.79 |
| | Edinburgh-LTG | 0.75 | 0.85 | 0.80 | 0.76 |
| | TRIOS | 0.80 | 0.74 | 0.77 | 0.77 |
| | JU_CSE | 0.48 | 0.56 | 0.52 | 0.53 |
| Spanish | TIPSem | 0.90 | 0.86 | 0.88 | 0.66 |
| | TIPSem-B | 0.92 | 0.85 | 0.88 | 0.66 |

Taula D.3: TempEval-2 - resultats de processament d'esdeveniments

| language | system | timex | dct | main | subord. |
|----------|--------------------|-------------|-------------|-------------|-------------|
| | | -event | -event | events | events |
| English | TRIOS ^a | 0.65 | 0.79 | 0.56 | 0.60 |
| | TIPSem | 0.55 | 0.82 | 0.55 | 0.59 |
| | TIPSem-B | 0.54 | 0.81 | 0.55 | 0.60 |
| | NCSU | 0.63 | 0.68 | 0.48 | 0.66 |
| | JU_CSE | 0.63 | 0.80 | 0.56 | 0.56 |
| | USFD2 | 0.63 | - | 0.45 | - |
| Spanish | TIPSem | 0.81 | 0.59 | - | - |
| | TIPSem-B | 0.81 | 0.59 | - | - |

^aTRIOS va ser l'únic que va deixar relacions sense categoritzar. No va ser penalitzat i per tant, els seus resultats no son estrictament comparables amb els altres.

Taula D.4: TempEval-2 - resultats de processament de relacions

Bibliografia

- Ahn, D., van Rantwijk, J., & de Rijke, M. (2007). A Cascaded Machine Learning Approach to Interpreting Temporal Expressions. In *NAACL*, pp. 420–427. ACL.
- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of ACM*, 26(11), 832–843.
- Allen, J. F. (1995). *Natural language understanding (2nd ed.)*. Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, EUA.
- Alonso, O., Berberich, K., Bedathur, S. J., & Weikum, G. (2010). NEAT: News Exploration Along Time. In *32nd European Conference on IR Research, ECIR*, Vol. 5993 of *LNCS*, p. 667.
- Alonso, O., Gertz, M., & Baeza-Yates, R. (2007a). On the Value of Temporal Information in Information Retrieval. *SIGIR Forum*, 41(2), 35–41.
- Alonso, O., Gertz, M., & Baeza-Yates, R. (2007b). Search Results using Timeline Visualizations. In *SIGIR '07: Proceedings of the 30th conference on Research and development in information retrieval*, pp. 908–908. ACM.
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9, 5–16.
- Beniger, J. R. (1986). *The control revolution: technological and economic origins of the information society*. Harvard University Press, Cambridge, MA, EUA.
- Berberich, K., Bedathur, S. J., Alonso, O., & Weikum, G. (2010). A language modeling approach for temporal information needs. In *32nd European Conference on IR Research, ECIR*, Vol. 5993 of *LNCS*, pp. 13–25.
- Bethard, S., & Martin, J. H. (2006). Identification of event mentions and their semantic class. In *EMNLP: Proceedings of the Conference on Empirical Methods in NLP*, pp. 146–154. ACL.

- Bethard, S., & Martin, J. H. (2007). CU-TMP: Temporal Relation Classification Using Syntactic and Semantic Features. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 129–132, Praga, Czech Republic. Association for Computational Linguistics.
- Bittar, A. (2009). Annotation of events and temporal expressions in French texts. In *ACL-IJCNLP '09*, pp. 48–51.
- Boguraev, B., & Ando, R. K. (2005). Effective Use of TimeBank for TimeML Analysis. In *Annotating, Extracting and Reasoning about Time and Events 05151*.
- Caselli, T. (2009). *Time, Events and Temporal Relations: an Empirical Model for Temporal Processing of Italian Texts*. Ph.D. thesis, Istituto Di Linguistica Computazionale, Consiglio Nazionale delle Ricerche, Pisa, Itàlia.
- Caselli, T., dell'Orletta, F., & Prodanof, I. (2009). TETI: a TimeML compliant TimEx tagger for Italian. In *Proceedings of the International Multiconference on Computer Science and Information Technology, IMCSIT*, pp. 185–192. IEEE.
- Caselli, T., Ide, N., & Bartolini, R. (2008). A bilingual corpus of inter-linked events. In (ELRA), E. L. R. A. (Ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pp. 2424–2429.
- Caselli, T., Llorens, H., Saquete, E., & Navarro, B. (2011). Data-Driven Approach Using Semantics for Recognizing and Classifying TimeML Events in Italian. In *RANLP. Submitted*.
- Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and maxent discriminative reranking. In *43rd Annual Meeting of the ACL*.
- Che, W., Li, Z., Li, Y., Guo, Y., Qin, B., & Liu, T. (2009). Multilingual dependency-based syntactic and semantic parsing. In *Proceedings of the 13th Conference on Computational Natural Language Learning: Shared Task, CoNLL '09*, pp. 49–54. ACL.
- Che, W., Li, Z., & Liu, T. (2010). LTP: A Chinese Language Technology Platform. In *Coling 2010: Demonstrations*, pp. 13–16, Beijing, Xina.
- Cheng, Y., Asahara, M., & Matsumoto, Y. (2007). NAIST.Japan: temporal relation identification using dependency parsed tree. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval)*, pp. 245–248. ACL.

- Comrie, B. (1985). *Tense*. Cambridge University Press, Cambridge.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Costa, F., & Branco, A. (2010). Temporal information processing of a new language: Fast porting with minimal resources. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 671–677, Uppsala, Suècia. ACL.
- Crystal, D. (2003). *A Dictionary of Linguistics and Phonetics (5th Ed.)*. Wiley-Blackwell.
- Cuenca, S., Grediaga, A., Llorens, H., & Albero, M. (2007a). Performance Evaluation of FPGA-Embedded Web Servers. In *Proceedings of the IEEE 14th International Conference in Electronics, Circuits and Systems (ICECS)*, pp. 1187–1190. IEEE.
- Cuenca, S., Llorens, H., & Albero, M. (2007b). Evaluación de servidores web embebidos sobre FPGAs. In *Actas de las VII Jornadas de Computación Reconfigurable y Aplicaciones (JCRA) del II Congreso Español de Informática (CEDI)*, pp. 221–228. Thomson.
- Cuenca, S., Ramos, H., Llorens, H., & Macia, P. (2008). Reconfigurable architecture for embedding web services. In *Proceedings of the IEEE 4th Southern Conference on Programmable Logic (SPL)*, pp. 119–124. IEEE.
- Daniel, N., Radev, D., & Allison, T. (2003). Sub-event based multi-document summarization. In *HLT-NAACL Text summarization workshop*, pp. 9–16. ACL.
- Derczynski, L., & Gaizauskas, R. (2010). USFD2: Annotating Temporal Expresions and TLINKs for TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 337–340, Uppsala, Suècia. Association for Computational Linguistics.
- Dowty, D. R. (1979). *Word Meaning and Montague Grammar*. Reidel, Dordrecht.
- Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics?. *Linguistics and Philosophy*, 9, 37–61.
- Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language*, 67(3).

- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. MIT Press.
- Ferro, L., Gerber, L., Mani, I., Sundheim, B., & Wilson, G. (2005). TIDES 2005 Standard for the Annotation of Temporal Expressions. Tech. rep., MITRE.
- Ferro, L., Mani, I., Sundheim, B., & Wilson, G. (2000). TIDES Temporal Annotation Guidelines, Draft v.1.0. Mitre technical report MTR00W0000094, MITRE.
- Filatova, E., & Hovy, E. (2001). Assigning time-stamps to event-clauses. In *Proceedings of the workshop on Temporal and spatial information processing*, pp. 1–8, Morristown, NJ, EUA. ACL.
- Fillmore, C. J. (1968). *Universals in linguistic theory*, chap. The Case for Case, pp. 1–88. Holt and Rinehart, and Winston, Nova York.
- Forascu, C. (2008). GMT to +2 or how can TimeML be used in Romanian. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pp. 3238–3242. ELRA.
- Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54, 199–227.
- Gaizauskas, R., & Wilks, Y. (1998). Information extraction: Beyond document retrieval. *Journal of Documentation*, 54(1), 70–105.
- Gerevini, A., Schubert, L., & Schaeffer, S. (1993). Temporal reasoning in Timegraph I–II. *SIGART Bulletin*, 4(3), 21–25.
- Gildea, D., & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28(3), 245–288.
- Grishman, R., & Sundheim, B. (1996). Message understanding conference-6: A brief history. In *COLING*, pp. 466–471.
- Gross, G., & Kiefer, F. (1995). La structure événementielle des substantifs. *Folia Linguistica*, 29(1-2), 45–65.
- Grover, C., Tobin, R., Alex, B., & Byrne, K. (2010). Edinburgh-LTG: TempEval-2 System Description. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 333–336, Uppsala, Suècia. Association for Computational Linguistics.
- Gruber, J. S. (1965). *Studies in lexical relations*. Ph.D. thesis, Massachusetts Institute of Technology.

- Ha, E., Baikadi, A., Licata, C., & Lester, J. (2010). NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 341–344, Uppsala, Suècia. Association for Computational Linguistics.
- Hacioglu, K., Chen, Y., & Douglas, B. (2005). Automatic Time Expression Labeling for English and Chinese Text. In *Computational Linguistics and Intelligent Text Processing*, Vol. 3406 of *LNCS*, pp. 548–559. Springer.
- Hagège, C., & Tannier, X. (2007). XRCE-T: XIP temporal module for TempEval campaign. In *TempEval (SemEval)*, pp. 492–495, Praga, Czech Republic. ACL.
- Hepple, M., Setzer, A., & Gaizauskas, R. (2007). USFD: preliminary exploration of features and classifiers for the TempEval-2007 tasks. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval)*, pp. 438–441. ACL.
- Hwang, C. H., & Schubert, L. K. (1992). Tense trees as the “fine structure” of discourse. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, pp. 232–240, Morristown, NJ, EUA. ACL.
- Im, S., You, H., Jang, H., Nam, S., & Shin, H. (2009). KTimeML: Specification of temporal and event expressions in Korean text. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 115–122. ACL.
- ISO-TimeML (2007). ISO/TC 37/SC 4/WG 2. Language Resource Management – Semantic Annotation Framework (SemAF) - Part 1: Time and events. Tech. rep., International Organization for Standardization.
- Johnson, C. R., Fillmore, C. J., Petruck, M. R. L., Baker, C. F., Ellsworth, M., Ruppenhofer, J., & Wood, E. J. (2002). FrameNet: Theory and Practice. Tech. rep., International Computer Science Institute.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall PTR, Upper Saddle River, NJ, EUA.
- Kamp, H., & Reyle, U. (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy. Dordrecht, Holanda.

- Katz, G., & Arosio, F. (2001). The annotation of temporal information in natural language sentences. In *Proceedings of the workshop on Temporal and spatial information processing*, Vol. 13, pp. 151–158, Morristown, NJ, EUA. ACL.
- Kiparsky, P. (2002). On the architecture of panini’s grammar. In *Three lectures delivered at the Hyderabad Conference on the architecture of grammar, Jan. 2002, and at UCLA March 2002*.
- Kolomiyets, O., & Moens, M.-F. (2010). Kul: Recognition and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 325–328, Uppsala, Suècia. Association for Computational Linguistics.
- Kolya, A. K., Ekbal, A., & Bandyopadhyay, S. (2010). Ju_cse_temp: A first step towards evaluating events, time expressions and temporal relations. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 345–350, Uppsala, Suècia. Association for Computational Linguistics.
- Lafferty, J. D., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th ICML*, pp. 282–289. Morgan Kaufmann.
- Lapata, M., & Lascarides, A. (2006). Learning sentence-internal temporal relations. *Journal of AI Research (JAIR)*, 27, 85–117.
- Lascarides, A., & Asher, N. (1993). A semantics and pragmatics for the pluperfect. In *Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics*, pp. 250–259, Morristown, NJ, EUA. ACL.
- Levin, B., & Hovav, M. R. (2005). *Argument Realization*. Research Surveys in Linguistics. Cambridge University Press.
- Li, W., Wong, K.-F., Cao, G., & Yuan, C. (2004). Applying machine learning to chinese temporal relation resolution. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL’04), Main Volume*, pp. 582–588, Barcelona.
- Li, W., Wong, K.-F., & Yuan, C. (2001). A model for processing temporal references in Chinese. In *Proceedings of the workshop on Temporal and Spatial Information Processing*, pp. 33–40. ACL.

- Llorens, H., Ivars, X., Buiges, A., Cabo, M., Climent, P., & Fuster, A. (2005). *DomoSim, Simulador de sistemas domóticos*, pp. 119–126. S.P. Universitat d’Alacant, Alacant.
- Llorens, H., Navarro, B., & Saquete, E. (2009a). Detección de Expresiones Temporales TimeML en Catalán mediante Roles Semánticos y Redes Semánticas. In *Procesamiento del Lenguaje Natural (SEPLN)*, 43, pp. 13–21.
- Llorens, H., Navarro, B., & Saquete, E. (2009b). From Semantic Roles to Temporal Information Representation. In *the 8th Mexican International Conference on Artificial Intelligence (MICA)*, Vol. 5845 of *LNAI*, pp. 124–135. LNCS.
- Llorens, H., Navarro, B., & Saquete, E. (2009c). Using Semantic Networks to Identify Temporal Expressions from Semantic Roles. In *VI RANLP*, pp. 219–224.
- Llorens, H., & Saquete, E. (2008). Búsqueda de respuestas basada en conocimiento semántico. Tech. rep., Universitat d’Alacant, Alacant.
- Llorens, H., Saquete, E., & Navarro, B. (2009). Temporal Expression Identification based on Semantic Roles. In *the 14th International Conference on Applications of Natural Language to Information Systems (NLDB)*, pp. 230–242. LNCS.
- Llorens, H., Saquete, E., & Navarro, B. (2011a). Syntax-motivated context windows of morpho-lexical features for learning time and event expression recognizers in natural language. In *NLDB*, LNCS.
- Llorens, H., Saquete, E., Navarro, B., & Gaizauskas, R. (2011b). Time-surfer: Time-based graphical access to document content. In *ECIR*, Vol. 6611 of *LNCS - Advances in Information Retrieval*, pp. 767–771. Springer.
- Llorens, H., Saquete, E., Navarro, B., Li, L., & He, Z. (2011c). Data-Driven Approach Based on Semantic Roles for Recognizing Temporal Expressions and Events in Chinese. In *NLDB*, LNCS.
- Llorens, H., Saquete, E., & Navarro-Colorado, B. (2010a). TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 725–733.
- Llorens, H., Saquete, E., & Navarro-Colorado, B. (2010b). TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2.

- In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 284–291. ACL.
- Llorens, H., Saquete, E., & Navarro-Colorado, B. (2011). Applying semantic knowledge to the automatic processing of temporal expressions and events in natural language. *Information Processing and Management (IPM)*. Submitted: Dec. 2010.
- Lloret, E., Llorens, H., Moreda, P., Saquete, E., & Palomar, M. (2011). Text summarization contribution to semantic question answering: New approaches for finding answers on the web. *Intelligent Journal of Intelligent Systems*. Submitted Jan 2011.
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge University Press.
- Lyons, J. (1981). *Language and linguistics: an introduction*. Cambridge University Press.
- Mani, I., Pustejovsky, J., & Gaizauskas, R. (Eds.). (2005). *The Language of Time: A Reader*. Oxford University Press.
- Mani, I., Verhagen, M., Wellner, B., Lee, C. M., & Pustejovsky, J. (2006). Machine learning of temporal relations. In *44th annual meeting of the ACL*, pp. 753–760, NJ, EUA. ACL.
- Mani, I., & Wilson, G. (2000). Robust temporal processing of news. In *ACL Annual Meeting*, pp. 69–76, NJ, EUA. ACL.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, EUA.
- March, O., & Baldwin, T. (2008). Automatic event reference identification. In *ALTA 2008*, pp. 79–87, Austràlia.
- McCallum, A., & Li, W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *HLT-NAACL*, pp. 188–191.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press.
- Melli, G., Z. Shi, Y. Wang, Y. L., & Popowich, F. (2006). Description of SQUASH, the SFU Question Answering Summary Handler for the DUC-2006 Summarization Task. In *DUC*.
- Min, C., Srikanth, M., & Fowler, A. (2007). Lcc-te: a hybrid approach to temporal relation identification in news text. In *Proceedings of the*

- 4th International Workshop on Semantic Evaluations (SemEval)*, pp. 219–222. ACL.
- Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, 14, 15–28.
- Morante, R. (2008). Semantic role labeling tools trained on the cast3lb-connl-semrol corpus. In *LREC*.
- Moreda, P. (2008). *Los Roles Semánticos en la Tecnología del Lenguaje Humano: Anotación y Aplicación*. Phd, Departament de Llenguatges i Sistemes Informàtics. Universitat d'Alacant.
- Moreda, P., Llorens, H., Saquete, E., & Palomar, M. (2008a). Automatic Generalization of a QA Answer Extraction Module Based on Semantic Roles. In *AAI - IBERAMIA*, Vol. 5290 of *LNAI, LNCS*, pp. 233–242. Springer.
- Moreda, P., Llorens, H., Saquete, E., & Palomar, M. (2008b). The influence of semantic roles in QA: a comparative analysis. In *Procesamiento del Lenguaje Natural (SEPLN)*, Vol. 41, pp. 55–62, Madrid.
- Moreda, P., Llorens, H., Saquete, E., & Palomar, M. (2008c). Two Proposals of a QA Answer Extraction Module Based on Semantic Roles. In *AAI - MICAI*, Vol. 5317 of *LNAI, LNCS*, pp. 174–184. Springer.
- Moreda, P., Llorens, H., Saquete, E., & Palomar, M. (2011). Combining semantic information in question answering systems. *Information Processing & Management, In Press, Corrected Proof*, –.
- Moreda, P., Navarro, B., & Palomar, M. (2007). Corpus-based semantic role approach in information retrieval. *Data Knowledge Engineering*, 61(3), 467–483.
- Moreno, L., Palomar, M., Molina, A., & Ferrández, A. (1999). *Introducción al Procesamiento del Lenguaje Natural*. Universitat d'Alacant.
- Moulin, B. (1997). Temporal contexts for discourse representation: An extension of the conceptual graph approach. *Applied Intelligence*, 7(3), 227–255.
- Navarro, B., Civit, M., Martí, M. A., Marcos, R., & Fernández, B. (2003). Syntactic, semantic and pragmatic annotation in Cast3LB. In *corpus linguistics (srolac)*.
- Negri, M., & Marseglia, L. (2004). Recognition and Normalization of Time Expressions: ITC-first at TERN 2004. Tech. rep., Information Society Technologies.

- Padró, L., Collado, M., Reese, S., Lloberes, M., & Castellón, I. (2010). Freeling 2.1: Five years of open-source language processing tools. In *Proceedings of 7th Language Resources and Evaluation Conference (LREC'10)*, La Valletta, Malta.
- Palmer, M., Gildea, D., & Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31.
- Pan, F., Mulkar, R., & Hobbs, J. R. (2006). Learning event durations from event descriptions. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 393–400. ACL.
- Pardiño, M., Gómez, J., Llorens, H., Muñoz-Terol, R., Navarro-Colorado, B., Saquete, E., Martínez-Barco, P., Moreda, P., & Palomar, M. (2009). IBQAst: A Question Answering System for Text Transcriptions. In *Evaluating Systems for Multilingual and Multimodal Information Access, CLEF 2008*, Vol. 5706, pp. 488–491. LNCS.
- Pardiño, M., Gómez, J. M., Llorens, H., Muñoz, R., Navarro, B., Saquete, E., Martínez-Barco, P., Moreda, P., & Palomar, M. (2008). Adapting IBQAS to work with text transcriptions in QAst Task: IBQAst. In *Cross Language Evaluation Forum (CLEF)*.
- Punyakanok, V., Roth, D., Yih, W., Zimak, D., & Tu, Y. (2004). Semantic role labeling via generalized inference over classifiers. In *HLT-NAACL (CoNLL)*, pp. 130–133. ACL.
- Puscasu, G. (2007). Wvali: Temporal relation identification by syntactico-semantic analysis. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 484–487, Praga, Czech Republic. Association for Computational Linguistics.
- Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.
- Pustejovsky, J. (2002). TERQAS: Time and Event Recognition for Question Answering Systems. In *ARDA Workshop*.
- Pustejovsky, J., Castaño, J. M., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., & Katz, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In *IWCS-5*.
- Pustejovsky, J., Gaizauskas, R., Saurí, R., Setzer, A., & Ingria, R. (2002). Annotation Guideline to TimeML I.0, available at <http://timeml.org>. Annotation specifications.

- Pustejovsky, J., Hanks, P., Saurí, R., See, A., Gaizauskas, R., Setzer, A., Radev, D. R., Sundheim, B., Day, D., Ferro, L., & Lazo, M. (2003). The TIMEBANK Corpus. In *Corpus Linguistics*, pp. 647–656.
- Reichenbach, H. (1947). *Elements of Symbolic Logic*, chap. The Tenses of Verbs, pp. 287–298. No. 51. The Macmillan Company, Nova York.
- Restrepo, F., Llorens, H., & Cuenca, S. (2008). Implementación en FPGA de una cámara IP. In *Actas de las Jornadas JCRA*. Thomson.
- Robaldo, L., Caselli, T., Russo, I., & Grella, M. (2011). From Italian Text To TimeML Document Via Dependency Parsing. In *Proceedings of the 12th CICLing*, pp. 177–187.
- Ruimy, N., Monachini, M., Gola, E., Calzolari, N., Fiorentino, M. C. D., Ulivieri, M., & Rossi, S. (2003). A computational semantic lexicon of Italian: SIMPLE. *Linguistica Computazionale, Computational Linguistics in Pisa, special Issue, XVIII-XIX*, 821–64.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2005). FrameNet II: Extended theory and practice. Tech. rep., International Computer Science Institute.
- Saquete, E. (2010). ID 392: TERSEO+T2T3 Transducer. A systems for Recognizing and Normalizing TIMEX3. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 317–320, Uppsala, Suècia. Association for Computational Linguistics.
- Saquete, E., González, J. L. V., Martínez-Barco, P., Muñoz, R., & Llorens, H. (2009). Enhancing QA Systems with Complex Temporal Question Processing Capabilities. *Journal of Artificial Intelligence Research (JAIR)*, 35, 775–811.
- Saquete, E., Martínez-Barco, P., & Muñoz, R. (2004). Automatic Multilinguality for Time Expression Resolution. In *MICAI*, Vol. 2972 of *LNCS*, pp. 458–467.
- Saquete, E., Muñoz, R., & Martínez-Barco, P. (2006). Event ordering using TERSEO system. *Data Knowledge Engineering*, 58(1), 70–89.
- Saurí, R. (2010). Tempeval 2. Spanish data release. Tech. rep. BM 2010-05, Barcelona Media Technical Report.
- Saurí, R., Knippen, R., Verhagen, M., & Pustejovsky, J. (2005). Evita: A Robust Event Recognizer For QA Systems. In *HLT/EMNLP. ACL*.

- Saurí, R., Saquete, E., & Pustejovsky, J. (2009). Annotating time expressions in Spanish. TimeML annotation guidelines. Tech. rep., Barcelona Media - Innovation Center.
- Schilder, F., & Habel, C. (2001). From temporal expressions to temporal information: semantic tagging of news messages. In *Proceedings of the workshop on Temporal and spatial information processing*, Vol. 13, pp. 91–98, Morristown, NJ, EUA. ACL.
- Schilder, F., Katz, G., & Pustejovsky, J. (2007). *Annotating, Extracting and Reasoning About Time and Events (Dagstuhl 2005)*, Vol. 4795 of *LNCs*. Springer.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pp. 44–49.
- Seker, S. E., & Diri, B. (2010). TimeML and Turkish Temporal Logic. In *IC-AI'10*, pp. 881–887.
- Setzer, A. (2001). *Temporal information in newswire articles: An annotation scheme and corpus study*. Ph.D. thesis, University of Sheffield.
- Setzer, A., & Gaizauskas, R. (2000). Annotating Events and Temporal Information in Newswire Texts. In *LREC 2000*, pp. 1287–1294.
- Sha, F., & Pereira, F. C. N. (2003). Shallow parsing with conditional random fields. In *HLT-NAACL*.
- Smith, C. (1991). *The Parameter of Aspect*. Kluwer, Dordrecht, Holanda.
- Song, F., & Cohen, R. (1991). Tense interpretation in the context of narrative. In *AAAI*, pp. 131–136.
- Strötgen, J., & Gertz, M. (2010). HeidelTime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 321–324, Uppsala, Suècia. Association for Computational Linguistics.
- Taulé, M., Martí, M. A., & Recasens, M. (2008). AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In ELRA (Ed.), *LREC*, Marràqueix, el Marroc.
- ter Meulen, A. (1995). *Representing time in natural language: the dynamic interpretation of tense and aspect*. Mit Press, Cambridge.
- TERN (2004). Time Expression Recognition and Normalization Evaluation Workshop (<http://focofoca.mitre.org/tern.html>). Conference.

- UzZaman, N., & Allen, J. F. (2010). TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 276–283, Uppsala, Suècia. Association for Computational Linguistics.
- Vendler, Z. (1967). *Linguistics and philosophy*, chap. Verbs and times, pp. 97–121. Cornell University Press, Ithaca, NY.
- Verhagen, M. (2007). Drawing TimeML Relations with TBox. In *Proceedings of the 2005 international conference on Annotating, extracting and reasoning about time and events*, pp. 7–28. Springer-Verlag.
- Verhagen, M., Gaizauskas, R., Hepple, M., Schilder, F., Katz, G., & Pustejovsky, J. (2007). Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 75–80, Praga. ACL.
- Verhagen, M., Mani, I., Saurí, R., Knippen, R., Jang, S. B., Littman, J., Rumshisky, A., Phillips, J., & Pustejovsky, J. (2005). Automating temporal annotation with TARSQI. In *ACL*, pp. 81–84, NJ, EUA. ACL.
- Verhagen, M., Saurí, R., Caselli, T., & Pustejovsky, J. (2010). Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62, Uppsala, Suècia. ACL.
- Vicente-Díez, M. T., Moreno-Schneider, J., & Martínez, P. (2010). UC3M system: Determining the extent, type and value of time expressions in TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 329–332, Uppsala, Suècia. Association for Computational Linguistics.
- Vilain, M. (1982). A system for reasoning about time. In *AAAI*, pp. 197–201.
- Vilain, M., Kautz, H., & van Beek, P. (1990). *Readings in Qualitative Reasoning about Physical Systems*, chap. Constraint propagation algorithms for temporal reasoning: a revised report, pp. 373–381. Morgan Kaufmann.
- Vossen, P. (1998). *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, MA, EUA.
- Webber, B. L. (1988). Tense as discourse anaphor. *Computational Linguistics*, 14, 61–73.
- Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. MIT Press.

- Wilson, G., Mani, I., Sundheim, B., & Ferro, L. (2001). A multilingual approach to annotating and extracting temporal information. In *Proceedings of the workshop on Temporal and Spatial information processing*, pp. 81–87, NJ, EUA. ACL.
- Xue, N. (2008). Labeling chinese predicates with semantic roles. *Computational Linguistics*, 34(2), 225–255.
- Xue, N., & Zhou, Y. (2010). Applying Syntactic, Semantic and Discourse Constraints in Chinese Temporal Annotation. In *Coling 2010: Posters*, pp. 1363–1372, Beijing, Xina.
- You, L., & Liu, K. (2005). Building Chinese FrameNet Database. In *Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE '05)*, pp. 301 – 306.
- Zhang, C., Cao, C., Niu, Z., & Yang, Q. (2008). A transformation-based error-driven learning approach for Chinese temporal information extraction. In *Proceedings of the 4th Asia information retrieval conference on Information retrieval technology (AIRS'08)*, pp. 663–669. Springer-Verlag.